

AK056137 Homo sapi
BD176860 A method
AL359588 Homo sapi
AX877611 Sequence
BD156750 Primer fo
AK001821 Homo sapi
BC006388 Mus muscu
AC063950 Homo sapi
AC005409 Homo sapi
BC071687 Homo sapi
CQ052066 Sequence
CQ087110 Sequence
CQ094163 Sequence
CQ200647 Sequence
CQ231806 Sequence
CQ328863 Sequence
CQ057085 Sequence
CQ076348 Sequence
CQ107339 Sequence
CQ205801 Sequence
CQ304361 Sequence
CQ341629 Sequence
L36434 Mus Musculu
AX806612 Sequence
AX869219 Sequence
BD149281 Primer fo

20 2587.4 46.2 2603 9 AK056137
21 2372.2 42.4 2428 6 BD176860
22 2372.2 42.4 2428 6 HSM802690
23 2018.2 36.1 2076 6 AX877611
24 2018.2 36.1 2076 6 BD156750
25 2018.2 36.1 2076 6 AK001821
26 1795.4 32.1 4699 10 BC006388
C 27 1716.4 30.7 177361 9 AC063950
C 28 1716.4 30.7 177361 9 AC005409
C 29 1544.2 27.6 1718 9 BC071687
C 30 1104.8 19.7 1962 6 CQ052066
C 31 1104.8 19.7 1962 6 CQ087110
C 32 1104.8 19.7 1962 6 CQ094163
C 33 1104.8 19.7 1962 6 CQ200647
C 34 1104.8 19.7 1962 6 CQ231806
C 35 1104.8 19.7 1962 6 CQ328863
C 36 1103.4 19.7 1125 6 CQ057085
C 37 1103.4 19.7 1125 6 CQ076348
C 38 1103.4 19.7 1125 6 CQ107339
C 39 1103.4 19.7 1125 6 CQ205801
C 40 1103.4 19.7 1125 6 CQ304361
C 41 1103.4 19.7 1125 6 CQ341629
C 42 869 15.5 1846 10 MUSBDL2
43 724.8 12.9 728 6 AX806612
44 692 12.4 780 6 AX869219
45 692 12.4 780 6 BD149281

ALIGNMENTS

RESULT 1
AR300462 AR300462 5597 bp DNA linear PAT 12-JUN-2003
LOCUS Sequence 4 from patent US 6537785.
DEFINITION AR300462
ACCESSION AR300462
VERSION AR300462.1 GI:31687904
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5597)
AUTHORS Canfield, W.M.
TITLE Methods of treating lysosomal storage diseases
JOURNAL Patent: US 6537785-A 4 25-MAR-2003,
FEATURES Location/Qualifiers
source
1..5597
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 5597; DB 6; Length 5597;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CGGAGCCGAGCGGGCGTCCGTCCGCGAGCTGCAATGAGCGCGCCCGAGGCTGTGACC 60
Db 1 CGGAGCCGAGCGGGCGTCCGTCCGCGAGCTGCAATGAGCGCGCCCGAGGCTGTGACC 60
Qy 61 TGC CGCGGGCGGGCCCGACCGGGCCCTGATGCGGGCTCGCTGAGCGCGCGGGCGGC 120
Db 61 TGC CGCGGGCGGGCCCGACCGGGCCCTGATGCGGGCTCGCTGAGCGCGCGGGCGGC 120
Qy 121 GCGGGCTCAGGCTCCCTCGGGCGGTGGCGGTGAAAGGGGTGATGCTGTTCAAGCTCC 180
Db 121 GCGGGCTCAGGCTCCCTCGGGCGGTGGCGGTGAAAGGGGTGATGCTGTTCAAGCTCC 180
Qy 181 TGCGAGACAAACCTATACCTGCTCTGCCACAGGTATGGGCTCTACGTGCTTCTTGG 240
Db 181 TGCGAGACAAACCTATACCTGCTCTGCCACAGGTATGGGCTCTACGTGCTTCTTGG 240
Qy 241 CGCTCGTGTGCACCATGCTCCGCTTCAGTTCCGAGAGGTGTTCTTGGAAATGAGCC 300

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: November 21, 2004, 20:12:11 ; Search time 22705 Seconds
(without alignments)
11657.364 Million cell updates/sec

Title: US-10-023-888-3
Perfect score: 5597
Sequence: 1 cggagccgagcggcgctccg.....aaaagttaattttgaaa 5597
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 4526729 seqs, 23644849745 residues
Total number of hits satisfying chosen parameters: 9053458
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5597	100.0	5597	6 AR300462	AR300462 Sequence
2	5597	100.0	5597	6 AR428739	AR428739 Sequence
3	5597	100.0	5597	6 AR442828	AR442828 Sequence
4	4475.2	80.0	4511	9 AB033034	AB033034 Homo sapi
5	4326	77.3	4369	6 AX468102	AX468102 Sequence
6	4057	72.5	4277	6 AX468106	AX468106 Sequence
7	3761	67.2	3783	6 AR300471	AR300471 Sequence
8	3761	67.2	3783	6 AR428748	AR428748 Sequence
9	3761	67.2	3783	6 AR442837	AR442837 Sequence
10	3485.4	62.3	3621	6 AR300472	AR300472 Sequence
11	3485.4	62.3	3621	6 AR428749	AR428749 Sequence
12	3485.4	62.3	3621	6 AR442838	AR442838 Sequence
13	3100.2	55.4	3150	9 BC042615	BC042615 Homo sapi
14	2874.6	51.4	5204	10 AK171312	AK171312 Mus muscu
15	2820.8	50.4	5229	6 AR300468	AR300468 Sequence
16	2820.8	50.4	5229	6 AR428745	AR428745 Sequence
17	2820.8	50.4	5229	6 AR442834	AR442834 Sequence
18	2755.8	49.2	3152	6 CQ119596	CQ119596 Sequence
19	2587.4	46.2	2603	6 AX1714069	AX1714069 Sequence

Db	241	CGCTCGTTGTCAACATCGTCTCCGCCCTCCAGTTCCGAGAGGTGGTTCTCGAATGGAGCC	300
Qy	301	GAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATTGCTGGAAAGTCCCTTC	360
Db	301		360
Qy	361	AGATCGGCTTTGTCGCCATGCGGATGACGTTGACGTTGTTTACACCTGGGTGAATGGCACAG	420
Db	361		420
Qy	421	ATCTTGAACACTGAGGAACTACAGCAGGTCAGAGACAGATGGAGAGGAGCAGAAAG	480
Db	421		480
Qy	481	CAATGAGAGAAATCCTTGGGAAACACAAACCGAACCTTACTAAGAGAGTGAGAACGAGT	540
Db	481		540
Qy	541	TAGAGTGTTCGTAACACACTGCAATTAAGTGGCAATGCTTGTAAGTGGAGCCGCTGC	600
Db	541		600
Qy	601	CAGCCAAACATCACCTGAGGAGCGTGCATCTCTTTATCTCTTTTCAATCTGCCAGTG	660
Db	601		660
Qy	661	ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCAACCAATGCTCAGTTGTTGTTTG	720
Db	661		720
Qy	721	ACAGTACTAAGAGTGTCAAGATGCCACTCTGACCTGCTTAAAGGAAATAGCAGACAGA	780
Db	721		780
Qy	781	CAGTATGAGGGGTACTTGACACAGATAAAGAGTCCCTGGATTAGTCTAATGCAAG	840
Db	781		840
Qy	841	ATTTGGCTTCTGAGTGATTTCCACCAATTCAGGAAACAACTCAACTTAAACAA	900
Db	841		900
Qy	901	AATTTGCCAGAAAATCTTTCTTAAAGTCAAACTGTTGCAAGTGTGAGTGTGAGGSCCAGTG	960
Db	901		960
Qy	961	TAGCGCTTCTAAACTGAAATACCCCAAGGATTTTCAAGAAATGAATGAACAACTAAGA	1020
Db	961		1020
Qy	1021	AGAACATCACATTTGAGGAAAGAACTGACCAATAGTCTGCAATTTTATTTGGGATC	1080
Db	1021		1080
Qy	1081	AGAACATCACATTTGAGGAAAGAACTGACCAATAGTCTGCAATTTTATTTGGGATC	1140
Db	1081		1140
Qy	1141	ACGAGAACTGAGGTACTTATGCGATCTATCGAGAGGATGACCACTGGGTTCCGGAATA	1200
Db	1141		1200
Qy	1201	TTTTCAATTTGACCAACGGGAGATTCATCTCGCTGGCTGAACTTGAACAACTCCCGAGTGA	1260
Db	1201		1260
Qy	1261	CAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTTACCTTTAGTTTAC	1320
Db	1261		1320
Qy	1321	CTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTGTCCCGAAGTCTTTATTTACCTAA	1380
Db	1321		1380
Qy	1381	ATGATGATGTCTATGTTTGGAAAGGATGTCTGCCAGATGATTTTTTACAGTCACTCCAAAG	1440
Db	1381		1440
Qy	1441	GCCAGAAGGTTTATTTGACATGCGCTGTGCCAACTGTGCCAGAGGCTGCCAGGTTCTCT	1500
Db	1441		1500
Qy	1501	GGATTAAAGGATCGCTATTGTGCAAGGCTTGTAAATTTAGCCCTGCGATTGGGATGGTG	1560
Db	1501		1560
Qy	1561	GGGATTGTCTCGAAACAGTGGAGGAGTGCCTATTATTGCGAGGAGTGGAGGTACTGGGA	1620
Db	1561		1620
Qy	1621	GTATTGGAGTGGACACCCCTGCGAGTTTGGTGGAGGAAATAACAGTGTCTCTTACTGTA	1680
Db	1621		1680
Qy	1681	ATCAGGGATGTGCGAAATTCCTGGCTCGCTGATAAGTTCTGTGAACCAAGCATGCAATGCT	1740
Db	1681		1740
Qy	1741	TGTCCTGTGGGTTTGTATGCTGGGACTCTGGGAGATCATTTTTCATGAATTTGATAAAG	1800
Db	1741		1800
Qy	1801	TGATCTCTTCCCAAAACAGACTCACTATATTATTTCAAAAGGTGAATGCTGCTTATT	1860
Db	1801		1860
Qy	1861	TCAGCTTTGCGAAGTAGCCTGAGGAGTGAAGTGGCTTATAGTGACAAATCCAAATAA	1920
Db	1861		1920
Qy	1921	TTTCGACATGCTTCTATTGCAACAGTGAAGGAAACCATCCCTCATATATGACAGTGA	1980
Db	1921		1980
Qy	1981	TGAATGCCACCAACAATATTTTAACTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	2040
Db	1981		2040
Qy	2041	AAATCCAGATACAGTGGAGTGGACACAGGAGGACCAAACTGAAATTTCTAGGCCCC	2100
Db	2041		2100
Qy	2101	AGAGGGTTTACGAAATTTAGTTAGTCCCAATAACACTTTTCCAGAGGCGGAAATCTCTT	2160
Db	2101		2160
Qy	2161	TTGAGGATATTTCCCAAGAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA	2220
Db	2161		2220
Qy	2221	CAAGGAGAGCCAGGAGAGGAGTGAATTTCCCTGTTAAATTTTCACTCTCTTCCAAAG	2280
Db	2221		2280
Qy	2281	AGCCCGAGTGGATGCTCAATACCTTGGATTTGCAATCTGAAACATGGAGACATCACTTGA	2340
Db	2281		2340
Qy	2341	AAGGATACAATTTGTCGAAGTCAAGTGTGCTGAGATCATTTCTGATGAATCACTCAGCATG	2400
Db	2341		2400
Qy	2401	CTAAATTAATAAATACTAATAAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	2460
Db	2401		2460

QY 841 ATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAACAA 900
DB 841 ATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAACAA 900
QY 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTGTATTCAGAGCCAGTG 960
DB 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTGTATTCAGAGCCAGTG 960
QY 961 TAGCCCTTTAAACCTGAATACCCCAAGGATTTTCAAGATTTGAATAAGCAAACTAAGA 1020
DB 961 TAGCCCTTTAAACCTGAATACCCCAAGGATTTTCAAGATTTGAATAAGCAAACTAAGA 1020
QY 1021 AGAACATGACCATTCATGAGAAAGAACTGACCAATAGTCTGCAATATTTATGGGATC 1080
DB 1021 AGAACATGACCATTCATGAGAAAGAACTGACCAATAGTCTGCAATATTTATGGGATC 1080
QY 1081 TGAGGCCATGAGCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140
DB 1081 TGAGGCCATGAGCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140
QY 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCGGAATA 1200
DB 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCGGAATA 1200
QY 1201 TTTTCATTTGACCAACGGGAGATTTCCATCTGCTGCTGAACCTTTGAACAATCTCGAGTGA 1260
DB 1201 TTTTCATTTGACCAACGGGAGATTTCCATCTGCTGCTGAACCTTTGAACAATCTCGAGTGA 1260
QY 1261 CAATAGTAAACACACAGAGTGTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320
DB 1261 CAATAGTAAACACACAGAGTGTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320
QY 1321 CTGCTATTGAAAGTCAACATTCATCGCATCGAAGGCTGTCCAGAGTTTATTTACCTAA 1380
DB 1321 CTGCTATTGAAAGTCAACATTCATCGCATCGAAGGCTGTCCAGAGTTTATTTACCTAA 1380
QY 1381 ATGATGATGATGTTTGGGAGGATGTCTGCCAGATGATTTTACAGTCACTCCAAAG 1440
DB 1381 ATGATGATGATGTTTGGGAGGATGTCTGCCAGATGATTTTACAGTCACTCCAAAG 1440
QY 1441 GCCAAGGTTTATTTGACATGGCTGTGCCAACTGTCCGAGGCTGCCAGGTTCT 1500
DB 1441 GCCAAGGTTTATTTGACATGGCTGTGCCAACTGTCCGAGGCTGCCAGGTTCT 1500
QY 1501 GGATTAAGATGGCTATTGTGCAAGGCTTGTAAATTCAGCTGCGATTCGGATGGT 1560
DB 1501 GGATTAAGATGGCTATTGTGCAAGGCTTGTAAATTCAGCTGCGATTCGGATGGT 1560
QY 1561 GGGATTTGCTGGAAACACAGTGGAGGAGTCCGTATATTGCAGGAGTGGAGTCTGGGA 1620
DB 1561 GGGATTTGCTGGAAACACAGTGGAGGAGTCCGTATATTGCAGGAGTGGAGTCTGGGA 1620
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAACAGTGTCTCTTACTGTA 1680
DB 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAACAGTGTCTCTTACTGTA 1680
QY 1681 ATCAGGGATGCGGAATTCCTGGCTGCTGATAGTTCTGTGACCAAGCATGCAATGCT 1740
DB 1681 ATCAGGGATGCGGAATTCCTGGCTGCTGATAGTTCTGTGACCAAGCATGCAATGCT 1740
QY 1741 TGCTCTGTGGGTTGATGCTGGGAGTGTGGGCAAGATCAITTTTCAATGATTTGATAAG 1800
DB 1741 TGCTCTGTGGGTTGATGCTGGGAGTGTGGGCAAGATCAITTTTCAATGATTTGATAAG 1800
QY 1801 TGAATCTTCTCCAAACAGACTCTATATTTATTTCCAAAGGTGAATGCTGCTTAT 1860
DB 1801 TGAATCTTCTCCAAACAGACTCTATATTTATTTCCAAAGGTGAATGCTGCTTAT 1860
QY 1861 TCAGCTTTTCAGAGTACGCAAGAGGAGTGTGAAGGTCCTATAGTGACCAATCCAAATA 1920
DB 1861 TCAGCTTTTCAGAGTACGCAAGAGGAGTGTGAAGGTCCTATAGTGACCAATCCAAATA 1920

QY 1921 TTGCAATGCTTCTATTGGCAACAAGTGGAAACCATCCACCTCATAATGCAAGTGGAA 1980
DB 1921 TTGCAATGCTTCTATTGGCAACAAGTGGAAACCATCCACCTCATAATGCAAGTGGAA 1980
QY 1981 TGAATGCCACCAACATATCTTTAATCTCAAGTTTCAAAATACAAACGATGAAGATTCA 2040
DB 1981 TGAATGCCACCAACATATCTTTAATCTCAAGTTTCAAAATACAAACGATGAAGATTCA 2040
QY 2041 AATGCAAGATAACAGTGGAGGTGGACAACAAGGAGGACCAAACTGAATTTCTACGGCCC 2100
DB 2041 AATGCAAGATAACAGTGGAGGTGGACAACAAGGAGGACCAAACTGAATTTCTACGGCCC 2100
QY 2101 AGAAGGTTACGAAAATTTAGTTAGTTCCTCAATACACTTTTCCAGAGGCGGAAATCTTT 2160
DB 2101 AGAAGGTTACGAAAATTTAGTTAGTTCCTCAATACACTTTTCCAGAGGCGGAAATCTTT 2160
QY 2161 TTGAGGATATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGACATGATTTAATCTCAA 2220
DB 2161 TTGAGGATATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGACATGATTTAATCTCAA 2220
QY 2221 CAAGGAGAGCCAGGAGGAGTGAATAATTTCCCTGTGTAATTTTCACTCTTTCCAAAG 2280
DB 2221 CAAGGAGAGCCAGGAGGAGTGAATAATTTCCCTGTGTAATTTTCACTCTTTCCAAAG 2280
QY 2281 ACSCCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACATGAGACATCCTTTGA 2340
DB 2281 ACSCCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACATGAGACATCCTTTGA 2340
QY 2341 AAGGATACATTTTGTCCAGTCAGCCTTGTGAGATCATTTCTGATGAACTCACAGCATG 2400
DB 2341 AAGGATACATTTTGTCCAGTCAGCCTTGTGAGATCATTTCTGATGAACTCACAGCATG 2400
QY 2401 CTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTTGTGGTCTCCAC 2460
DB 2401 CTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTTGTGGTCTCCAC 2460
QY 2461 AGGAAACACAGTTTCAATAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGATTCG 2520
DB 2461 AGGAAACACAGTTTCAATAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGATTCG 2520
QY 2521 AGAGGTTGACTTTTCTGCACTGAGTGAATAAGTGAATGGTCAATGACAGGTCAGAAATC 2580
DB 2521 AGAGGTTGACTTTTCTGCACTGAGTGAATAAGTGAATGGTCAATGACAGGTCAGAAATC 2580
QY 2581 CACCCCTGGATTTGGAGACCAACAGCAAGATTTAGAGTGAAACTCACACCCAAACCA 2640
DB 2581 CACCCCTGGATTTGGAGACCAACAGCAAGATTTAGAGTGAAACTCACACCCAAACCA 2640
QY 2641 TAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700
DB 2641 TAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700
QY 2701 TGACAAAGAAAGAAATTCACAGGAAAGAAAGAAAGACAGTAGAATGGAGGAAATG 2760
DB 2701 TGACAAAGAAAGAAATTCACAGGAAAGAAAGAAAGACAGTAGAATGGAGGAAATG 2760
QY 2761 CTGAAATACATAGGCTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGATTACACAG 2820
DB 2761 CTGAAATACATAGGCTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGATTACACAG 2820
QY 2821 ATAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAAGTATTTTCAAGATCTTCTGAGC 2880
DB 2821 ATAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAAGTATTTTCAAGATCTTCTGAGC 2880
QY 2881 AAGAGAGTCAATTTGAAGACCAATTTGGCATCTTCACTGATAGCAAAATCTGAGGAGC 2940
DB 2881 AAGAGAGTCAATTTGAAGACCAATTTGGCATCTTCACTGATAGCAAAATCTGAGGAGC 2940
QY 2941 AACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000
DB 2941 AACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000
QY 3001 TTGATTTACATCGCGAAGAGTCCCTGCTCACATGCTCACATGATGACCGGATTTGTTA 3060

Db	3001	TTGGATTACATCGCGGGAAGTCCCTGCTCAGATGCCTCAGATGTTGACCGGATTGTTA	3060
Qy	3061	TGCRAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTTACAAAATGTC	3120
Db	3061	TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTTACAAAATGTC	3120
Qy	3121	GCATTCTGAGGATATGAGTTGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC	3180
Db	3121	GCATTCTGAGGATATGAGTTGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC	3180
Qy	3181	AGCCTAGTATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT	3240
Db	3181	AGCCTAGTATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT	3240
Qy	3241	CTGACAGAGAAATCCGAACACTGGCTACCGAATTCAGAACTGCCGTTAAGTTTGCGAGG	3300
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACCGAATTCAGAACTGCCGTTAAGTTTGCGAGG	3300
Qy	3301	ATTGACAGGCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360
Db	3301	ATTGACAGGCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360
Qy	3361	GCAGCTAAATAATATCCCACTCAGGAATCTCTATGATCCCAACTGCCACCGG	3420
Db	3361	GCAGCTAAATAATATCCCACTCAGGAATCTCTATGATCCCAACTGCCACCGG	3420
Qy	3421	TCATTAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCCAAGACATATA	3480
Db	3421	TCATTAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCCAAGACATATA	3480
Qy	3481	AGGACAAAACAAATATAGTTTGAATCATGCGAGAGAGAAATCGCTTTTAAATGA	3540
Db	3481	AGGACAAAACAAATATAGTTTGAATCATGCGAGAGAGAAATCGCTTTTAAATGA	3540
Qy	3541	TTCTGACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAACCACTAGGA	3600
Db	3541	TTCTGACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAACCACTAGGA	3600
Qy	3601	AGTTTGTGCTGATGACAACTGACCAATGCTAAGATGCTCAGACAGTGAAG	3660
Db	3601	AGTTTGTGCTGATGACAACTGACCAATGCTAAGATGCTCAGACAGTGAAG	3660
Qy	3661	CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCACTACCTTCCCAATTTGAACGTC	3720
Db	3661	CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCACTACCTTCCCAATTTGAACGTC	3720
Qy	3721	GAGATATCGAAACCGTTTCTCATATGATGAGCTGCGAGAAATGAGGCGCTTATCGAG	3780
Db	3721	GAGATATCGAAACCGTTTCTCATATGATGAGCTGCGAGAAATGAGGCGCTTATCGAG	3780
Qy	3781	ACAAATTCGAAGTTTGGACCCATGCTGATGACAACTGATGATGTTTACTATATCT	3840
Db	3781	ACAAATTCGAAGTTTGGACCCATGCTGATGACAACTGATGATGTTTACTATATCT	3840
Qy	3841	CATTTTGTGCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATTTCCAGAGGAGGATAC	3900
Db	3841	CATTTTGTGCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATTTCCAGAGGAGGATAC	3900
Qy	3901	ACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAACCACTTA	3960
Db	3901	ACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAACCACTTA	3960
Qy	3961	CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTTGTGATGAT	4020
Db	3961	CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTTGTGATGAT	4020
Qy	4021	GCTTAGCAGTTTGGCCGAGAGGAAATATCAGTACCATGCTGTTTGTGCGATGAA	4080
Db	4021	GCTTAGCAGTTTGGCCGAGAGGAAATATCAGTACCATGCTGTTTGTGCGATGAA	4080
Qy	4081	TATAGCCCACTGATAGGAAATTTTAACTAACCAACCACTGTAAGAACTGCTGTCAGCAGC	4140
Db	4081	TATAGCCCACTGATAGGAAATTTTAACTAACCAACCACTGTAAGAACTGCTGTCAGCAGC	4140

Db	4081	TATAGCCCACTGATAGGAAATTTTAACTAACCAACCACTGTAAGAACTGCTGTCAGCAGC	4140
Qy	4141	TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGTCATCTCTTTTATAAAAA	4200
Db	4141	TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGTCATCTCTTTTATAAAAA	4200
Qy	4201	GGCTCATCTCAAGAGACAGCTGTTAAATTTCCCAAGCAATCATTTGAGACTAATCTTAT	4260
Db	4201	GGCTCATCTCAAGAGACAGCTGTTAAATTTCCCAAGCAATCATTTGAGACTAATCTTAT	4260
Qy	4261	TAGGAGAGCCCTATGCCAGCTGGAGTATCTTAAGAGGCTCCAGTCTTTGCACTCCAA	4320
Db	4261	TAGGAGAGCCCTATGCCAGCTGGAGTATCTTAAGAGGCTCCAGTCTTTGCACTCCAA	4320
Qy	4321	AGCCTTTGCTAAAGTTTGCACCTTTTTTTTTTTCAATTTCCCAATTTTAAAGTAGTACTA	4380
Db	4321	AGCCTTTGCTAAAGTTTGCACCTTTTTTTTTTTCAATTTCCCAATTTTAAAGTAGTACTA	4380
Qy	4381	AGTTAACTAGTTTATCTTCTGCTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTA	4440
Db	4381	AGTTAACTAGTTTATCTTCTGCTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTA	4440
Qy	4441	TAGATGTTTATTTAAATAATGACGAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
Db	4441	TAGATGTTTATTTAAATAATGACGAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
Qy	4501	TTTTTATTAATTTAAGACTGTAAATGCTTTAAACCACTAACTGAGAGCTCAATG	4560
Db	4501	TTTTTATTAATTTAAGACTGTAAATGCTTTAAACCACTAACTGAGAGCTCAATG	4560
Qy	4561	ATTGACATCTGAATGCTTTGTAATTTGACTTCAGGCCCTTAAAGATGCTATGATTTCA	4620
Db	4561	ATTGACATCTGAATGCTTTGTAATTTGACTTCAGGCCCTTAAAGATGCTATGATTTCA	4620
Qy	4621	CGTGACGCTCTAATTTCAACAGGCTAGATGATGATCTACTTACGAGTGTAAATATGTTT	4680
Db	4621	CGTGACGCTCTAATTTCAACAGGCTAGATGATGATCTACTTACGAGTGTAAATATGTTT	4680
Qy	4681	TGGAATGTPACATATTTCAACAGAGAGTGCCTCATTTTGAAGATGATGCTGATGCA	4740
Db	4681	TGGAATGTPACATATTTCAACAGAGAGTGCCTCATTTTGAAGATGATGCTGATGCA	4740
Qy	4741	CTGGCACATTAACAGTGTGCTTGTTTAATACTCATTTGATATATTTCCAGTAGCTATCT	4800
Db	4741	CTGGCACATTAACAGTGTGCTTGTTTAATACTCATTTGATATATTTCCAGTAGCTATCT	4800
Qy	4801	CTCAGTTGGTTTTGATAGACAGAGGCGCAACATTTCTTTGTAAGGCTGTTAGT	4860
Db	4801	CTCAGTTGGTTTTGATAGACAGAGGCGCAACATTTCTTTGTAAGGCTGTTAGT	4860
Qy	4861	AAATTTATTCAGGCGCACCTGCTCTTTGTCATACATTTCTTGTGTTTGTGTTTGT	4920
Db	4861	AAATTTATTCAGGCGCACCTGCTCTTTGTCATACATTTCTTGTGTTTGTGTTTGT	4920
Qy	4921	TTTTTTTCAACAAACCTCTTAAAGATGTAAGACCATGTTTAGCTTGAGCTGTACAAA	4980
Db	4921	TTTTTTTCAACAAACCTCTTAAAGATGTAAGACCATGTTTAGCTTGAGCTGTACAAA	4980
Qy	4981	ACTGCCACACAGCAGATGTGACCTCAGGCGCATCATTTGCCAATCACTGAGAAATATT	5040
Db	4981	ACTGCCACACAGCAGATGTGACCTCAGGCGCATCATTTGCCAATCACTGAGAAATATT	5040
Qy	5041	TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	5100
Db	5041	TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	5100
Qy	5101	TGAGTGGGCGCAATCTCAGCTCAGTCAACCTCCCGGTTTCAAGCAGTCTCTGTC	5160
Db	5101	TGAGTGGGCGCAATCTCAGCTCAGTCAACCTCCCGGTTTCAAGCAGTCTCTGTC	5160
Qy	5161	TCAGGCTTTCTGAGTAGCTGGGACTACAGGTGATGCCACCAACCTGCTAAATTTTGT	5220
Db	5161	TCAGGCTTTCTGAGTAGCTGGGACTACAGGTGATGCCACCAACCTGCTAAATTTTGT	5220

QY 5221 TTTTATTAGACAGCGGGTCCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280
Db 5221 TTTTATTAGACAGCGGGTCCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280
QY 5281 AGGTGATCCACCTGCTCTCCCTCCAAAGTCTCAGATTAACAGGATTAAGCCAGTGCAC 5340
Db 5281 AGGTGATCCACCTGCTCTCCCTCCAAAGTCTCAGATTAACAGGATTAAGCCAGTGCAC 5340
QY 5341 CCAGCGGAGAAATAGTATTTTATGTATGGTTAAACCTTGGGCTAGCCATATTTATG 5400
Db 5341 CCAGCGGAGAAATAGTATTTTATGTATGGTTAAACCTTGGGCTAGCCATATTTATG 5400
QY 5401 TCATAATACATGGATTTGTGAAGAGCAGATTCATGAGTAACCTCTGACAGGATTTTATG 5460
Db 5401 TCATAATACATGGATTTGTGAAGAGCAGATTCATGAGTAACCTCTGACAGGATTTTATG 5460
QY 5461 ATCATGATCTCAACATATTTCTCCAAATGGCATACATCTTTGTACAAAGAACTTGA 5520
Db 5461 ATCATGATCTCAACATATTTCTCCAAATGGCATACATCTTTGTACAAAGAACTTGA 5520
QY 5521 ATGTAATACCTGTTGTGCTGAAGAGTGTGATTTTCAAACTGAACTCATATA 5580
Db 5521 ATGTAATACCTGTTGTGCTGAAGAGTGTGATTTTCAAACTGAACTCATATA 5580
QY 5581 AAGTTAAATTTTGA 5597
Db 5581 AAGTTAAATTTTGA 5597

RESULT 3
AR442828
LOCUS AR442828 5597 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 4 from patent US 6670165.
ACCESSION AR442828
VERSION AR442828.1 GI:42670304
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5597)
AUTHORS Canfield, W.M.
TITLE Methods for producing highly phosphorylated lysosomal hydrolases
JOURNAL Patent: US 6670165-A 4 30-DEC-2003;
FEATURES
Location/Qualifiers
1..5597
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 5597; DB 6; Length 5597;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGAGCGAGCGGGGCTCGCTCGCGGAGCTGCAATGAGCGCGCGCGGAGGCTGTGACC 60
Db 1 CGGAGCGAGCGGGGCTCGCTCGCGGAGCTGCAATGAGCGCGCGCGGAGGCTGTGACC 60
QY 61 TCG 120
Db 61 TCG 120
QY 121 GCGGCTCAGGCTCCTCGGCGGCTGCGGCTGCGGCTGAGAGGGGTGATGCTGTCAAGTCC 180
Db 121 GCGGCTCAGGCTCCTCGGCGGCTGCGGCTGCGGCTGAGAGGGGTGATGCTGTCAAGTCC 180
QY 181 TGCAGAGACAAACCTATACCTGCTGCTCCAGGATGAGGCTCTACGCTGCTCTTTGG 240
Db 181 TGCAGAGACAAACCTATACCTGCTGCTCCAGGATGAGGCTCTACGCTGCTCTTTGG 240
QY 241 GCGTGTGTACCATCTCTCCGCTTCCAGTTCGAGAGGTGGTCTCGAATGGAGCC 300
Db 241 GCGTGTGTACCATCTCTCCGCTTCCAGTTCGAGAGGTGGTCTCGAATGGAGCC 300

QY 301 GAGATCAATACCATGTTTGTGATTTCTATAGACAAATATTTGCTGGAAAGTCTCTTC 360
Db 301 GAGATCAATACCATGTTTGTGATTTCTATAGACAAATATTTGCTGGAAAGTCTCTTC 360
QY 361 AGAATCGGCTTTGCTGCGCCATCGCATTTGACGTTTGTACACCTGGGTGAATGGACAG 420
Db 361 AGAATCGGCTTTGCTGCGCCATCGCATTTGACGTTTGTACACCTGGGTGAATGGACAG 420
QY 421 ATCTTGAATCTATGAAGGAATCTACAGAGTCTAGAGAACAGATGAGAGGAGGAGAGAG 480
Db 421 ATCTTGAATCTATGAAGGAATCTACAGAGTCTAGAGAACAGATGAGAGGAGGAGAGAG 480
QY 481 CAATGAGAGAAATCTCTTGGGAAACACACAGCACTCTACTAGAGAGAGTGAAGACAGT 540
Db 481 CAATGAGAGAAATCTCTTGGGAAACACACAGCACTCTACTAGAGAGAGTGAAGACAGT 540
QY 541 TAGAGTGTTTGCTTAACACACTGCAATTAAGGTGCCAATGCTTGTACTGGACCCAGCCTGC 600
Db 541 TAGAGTGTTTGCTTAACACACTGCAATTAAGGTGCCAATGCTTGTACTGGACCCAGCCTGC 600
QY 601 CAGCCACATCAGCCTGAGGAGCTGCCATCTCTTTATCCTTCTTTTCTTCTTCTTCTTCT 660
Db 601 CAGCCACATCAGCCTGAGGAGCTGCCATCTCTTTATCCTTCTTTTCTTCTTCTTCTTCT 660
QY 661 ACATTTTCAATGTTGCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCA 720
Db 661 ACATTTTCAATGTTGCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCA 720
QY 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGAAATAGCAGACAGA 780
Db 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGAAATAGCAGACAGA 780
QY 781 CAGTATGGAGGGGTACTTGAACACAGATTAAGAGTCCCTGGATTTAGTCTTAATGCAAG 840
Db 781 CAGTATGGAGGGGTACTTGAACACAGATTAAGAGTCCCTGGATTTAGTCTTAATGCAAG 840
QY 841 ATTGCGTTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCAACTAAAAACAA 900
Db 841 ATTGCGTTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCAACTAAAAACAA 900
QY 901 AATTGCCAGAAATCTTCTCTTAAAGTCAAACTGTGAGTTGTGATTTAGAGCCAGTG 960
Db 901 AATTGCCAGAAATCTTCTCTTAAAGTCAAACTGTGAGTTGTGATTTAGAGCCAGTG 960
QY 961 TAGCGCTTCTTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAGA 1020
Db 961 TAGCGCTTCTTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAGA 1020
QY 1021 AGAATATGACCATTTGAGAAAGAACTGACCAATAGTCTTGCATTTTATTTATGGATC 1080
Db 1021 AGAATATGACCATTTGAGAAAGAACTGACCAATAGTCTTGCATTTTATTTATGGATC 1080
QY 1081 TGAGCGCATTCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140
Db 1081 TGAGCGCATTCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140
QY 1141 ACAGAGATCTGAGGTAATCTATGCGATCTATGAGAGGATGACCATGGTTTCGGAATA 1200
Db 1141 ACAGAGATCTGAGGTAATCTATGCGATCTATGAGAGGATGACCATGGTTTCGGAATA 1200
QY 1201 TTTTCTATGTCACCAACGGGAGATTTCCATCTCTGGCTGAACCTTGAACCTCTCGAGTGA 1260
Db 1201 TTTTCTATGTCACCAACGGGAGATTTCCATCTCTGGCTGAACCTTGAACCTCTCGAGTGA 1260
QY 1261 CAATAGTAACACACAGGATTTTTCGAAATTTGAGCCACTTGCCTACTTTAGTTTAC 1320
Db 1261 CAATAGTAACACACAGGATTTTTCGAAATTTGAGCCACTTGCCTACTTTAGTTTAC 1320
QY 1321 CTGCTATGGAAGTCACTTCATCGATCGAGGGCTGTCCAGAGAGTTTATTTACCTAA 1380
Db 1321 CTGCTATGGAAGTCACTTCATCGATCGAGGGCTGTCCAGAGAGTTTATTTACCTAA 1380

1381 ATGATGATGTCATGTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
1381 ATGATGATGTCATGTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
1441 GCCAAGAGGTTTATTTGACATGGCTGTGCGCAACTGTGCGAGGGCTGCCAGGTTTCCT 1500
1441 GCCAAGAGGTTTATTTGACATGGCTGTGCGCAACTGTGCGAGGGCTGCCAGGTTTCCT 1500
1501 GGATTAAGATGGCTTATTTGACAAAGCTTGTATAAATTCAGCCTCGCATGGATGGTG 1560
1501 GGATTAAGATGGCTTATTTGACAAAGCTTGTATAAATTCAGCCTCGCATGGATGGTG 1560
1561 GGGATGCTCTGGAAACAGTGGAGGAGTCCCTATATTGACAGAGGTGGAGTACTGGGA 1620
1561 GGGATGCTCTGGAAACAGTGGAGGAGTCCCTATATTGACAGAGGTGGAGTACTGGGA 1620
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAACAGTGTCTTACTGTA 1680
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAACAGTGTCTTACTGTA 1680
1681 ATCAGGAGTGTGGAAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740
1681 ATCAGGAGTGTGGAAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740
1741 TGTCTGTGGGTTTGATCTGGCAGCTGTGGCAAGATCATTTTCATGAATTTGTATAAAG 1800
1741 TGTCTGTGGGTTTGATCTGGCAGCTGTGGCAAGATCATTTTCATGAATTTGTATAAAG 1800
1801 TGATCTCTCCAAACAGACTCATATATATTTCBAAAGGTGAATGCGCTTATT 1860
1801 TGATCTCTCCAAACAGACTCATATATATTTCBAAAGGTGAATGCGCTTATT 1860
1861 TCAGCTTTGCAAGAGTACCAAAAGAGAGTGTGAAGGTGCTATAGTACAATCCATAA 1920
1861 TCAGCTTTGCAAGAGTACCAAAAGAGAGTGTGAAGGTGCTATAGTACAATCCATAA 1920
1921 TTCGATGCTCTTATTTGCCAAAGTGGAAAAACATCCACCTCATATGACAGTGGAA 1980
1921 TTCGATGCTCTTATTTGCCAAAGTGGAAAAACATCCACCTCATATGACAGTGGAA 1980
1981 TGAATGCCACACATATATTTTAACTCTCAGTTCCTCAAAATACAAAGTGAAGTTCA 2040
1981 TGAATGCCACACATATATTTTAACTCTCAGTTCCTCAAAATACAAAGTGAAGTTCA 2040
2041 AAATCAGATTAACAGTGGAGTGGACACAGGAGGAGGACCAAAACTGAAATTTACGGCCC 2100
2041 AAATCAGATTAACAGTGGAGTGGACACAGGAGGAGGACCAAAACTGAAATTTACGGCCC 2100
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACATCTTCTCAGAGCGGAAATCCTTT 2160
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACATCTTCTCAGAGCGGAAATCCTTT 2160
2161 TTGAGATATTTCCCAAGAAAAACGCTTCCGAAAGTTTAAAGACATGATTTAACTCAA 2220
2161 TTGAGATATTTCCCAAGAAAAACGCTTCCGAAAGTTTAAAGACATGATTTAACTCAA 2220
2221 CAAGAGAGCCAGGAAGAGTGAATAATTCCTCTGGTAAATATTTTCACTCTCCAAAG 2280
2221 CAAGAGAGCCAGGAAGAGTGAATAATTCCTCTGGTAAATATTTTCACTCTCCAAAG 2280
2281 AGCCCGAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGAGACATCACTTTGA 2340
2281 AGCCCGAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGAGACATCACTTTGA 2340
2341 AAGGATCAATTTGTCCAAAGTCAAGCTTGTGAGATCATTTTCAATGAACTCAGCATG 2400
2341 AAGGATCAATTTGTCCAAAGTCAAGCTTGTGAGATCATTTTCAATGAACTCAGCATG 2400
2401 CTAAATTAATAATCAAGCTTAAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460
2401 CTAAATTAATAATCAAGCTTAAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460
2461 AGGAAAAACAGGTTTCATATAAAGCATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520

2461 AGGAAAAACAGGTTTCATATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
2521 AGAGGTTGACCTTTTCTCGAGTGTGTAAAAGTGAATGGTCAACAGGCTCAGAATC 2580
2521 AGAGGTTGACCTTTTCTCGAGTGTGTAAAAGTGAATGGTCAACAGGCTCAGAATC 2580
2581 CACCCCTGGAATTCAGACACACAGCAAGATTTAGAGTGGAACTCACACCCAAAGAACCA 2640
2581 CACCCCTGGAATTCAGACACACAGCAAGATTTAGAGTGGAACTCACACCCAAAGAACCA 2640
2641 TAGGCGGAAATTCGAAAAAGAAACCCCATCTCTGATTGTTTCCACTGGAAAGCCAGA 2700
2641 TAGGCGGAAATTCGAAAAAGAAACCCCATCTCTGATTGTTTCCACTGGAAAGCCAGA 2700
2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAACAGTAGAATGAGGAAATG 2760
2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAACAGTAGAATGAGGAAATG 2760
2761 CTGAAATTCACATAGCGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTACACAG 2820
2761 CTGAAATTCACATAGCGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTACACAG 2820
2821 ATAGTTACTTGGCTTTTGGCATGGAGAAAGAAAGTATTTCCAGATCTTCTCGACG 2880
2821 ATAGTTACTTGGCTTTTGGCATGGAGAAAGAAAGTATTTCCAGATCTTCTCGACG 2880
2881 AAGAAGAGTCATTGGAAGACACAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC 2940
2881 AAGAAGAGTCATTGGAAGACACAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC 2940
2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAATAGCAAGT 3000
2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAATAGCAAGT 3000
3001 TTGGATTCACATCGCGGAAGTCCCTGCTCAGTCCCTCAGATATGTAATAAATTTCTAATAGCAAGT 3060
3001 TTGGATTCACATCGCGGAAGTCCCTGCTCAGTCCCTCAGATATGTAATAAATTTCTAATAGCAAGT 3060
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGACAGCTCATTTTCAAAAGTGC 3120
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGACAGCTCATTTTCAAAAGTGC 3120
3121 GCATTTGAGGATATGAGTTTGGCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180
3121 GCATTTGAGGATATGAGTTTGGCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180
3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
3241 CTGACAGAAATCCGAACTGCTACAGATTTCAAGATTCAGAACTGCGGTTAAGTTTGCAGG 3300
3241 CTGACAGAAATCCGAACTGCTGCTACAGATTTCAAGATTTCAAGATTTAAGTTTGCAGG 3300
3301 ATTTGACAGCTGCGAAACACATGCTAAATTAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
3301 ATTTGACAGCTGCGAAACACATGCTAAATTAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
3361 CGCAGCTAAATATTTCCCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420
3361 CGCAGCTAAATATTTCCCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420
3421 TCATTAAGAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATA 3480
3421 TCATTAAGAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATA 3480
3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
3541 TTCGTACCAAGCTTCTCATGTGGTTGGCCAGTTGGATGATTAAGAAAGAAACCTCAGGA 3600

Db 3541 TTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATGAAGAAAAACCCCTAGGA 3600
Qy 3601 AGTTTGTGTCCTGAATGACAAACATGTGACCAACATCAATAAGATGCTCAGACAGTGAAG 3660
Db 3601 AGTTTGTGTCCTGAATGACAAACATGTGACCAACATCAATAAGATGCTCAGACAGTGAAG 3660
Qy 3661 CTGTTCTCAGGAGCTTCATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720
Db 3661 CTGTTCTCAGGAGCTTCATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720
Qy 3721 GAGATATCGAAACCGTTTCCTTCATATGCAATGAGCTCGAGAAATGAGGCGCTTATCGAG 3780
Db 3721 GAGATATCGAAACCGTTTCCTTCATATGCAATGAGCTCGAGAAATGAGGCGCTTATCGAG 3780
Qy 3781 ACAAAATCGAAGTTTGGACCAATGCTGTACTAGCAACATGATATGTTTACTATATCT 3840
Db 3781 ACAAAATCGAAGTTTGGACCAATGCTGTACTAGCAACATGATATGTTTACTATATCT 3840
Qy 3841 CATTTTTTCTCAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAGAGGAGTAC 3900
Db 3841 CATTTTTTCTCAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAGAGGAGTAC 3900
Qy 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAATCTTCAATTTGAAACCACTCTA 3960
Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAATCTTCAATTTGAAACCACTCTA 3960
Qy 3961 CCTCAGCATTTACTCAGCATTTTAAACTCAGTTCACAGAGATGCTCTTGTGTGATGAT 4020
Db 3961 CCTCAGCATTTACTCAGCATTTTAAACTCAGTTCACAGAGATGCTCTTGTGTGATGAT 4020
Qy 4021 GCTTAGCAGTTTGGCCCGAAGAAATATCCAGTACCAATGCTGTTTGTGCAATGAA 4080
Db 4021 GCTTAGCAGTTTGGCCCGAAGAAATATCCAGTACCAATGCTGTTTGTGCAATGAA 4080
Qy 4081 TATAGCCCACTGACTAGCAATTTTAAACCAACCACTGAAACCTTGTGTCGAGGAGC 4140
Db 4081 TATAGCCCACTGACTAGCAATTTTAAACCAACCACTGAAACCTTGTGTCGAGGAGC 4140
Qy 4141 TCTGAACCTGATTTTAAAGAAATTTGCTATGACCTGCTCATCTCTTTTATAAAA 4200
Db 4141 TCTGAACCTGATTTTAAAGAAATTTGCTATGACCTGCTCATCTCTTTTATAAAA 4200
Qy 4201 GCCTCAGTGAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGACACTTAT 4260
Db 4201 GCCTCAGTGAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGACACTTAT 4260
Qy 4261 TAGGAGAGCTTATCCAGCTGGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320
Db 4261 TAGGAGAGCTTATCCAGCTGGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320
Qy 4321 AGCCTTTTCTAAAGTTTGCATTTTCTTTTTCATTTCCCATTTTAAAGTAGTACTA 4380
Db 4321 AGCCTTTTCTAAAGTTTGCATTTTCTTTTTCATTTCCCATTTTAAAGTAGTACTA 4380
Qy 4381 AGTTAACTAGTTATTTCTGCTTCAGTATACGAATTTGGGATGCTTAAACCTATTTTA 4440
Db 4381 AGTTAACTAGTTATTTCTGCTTCAGTATACGAATTTGGGATGCTTAAACCTATTTTA 4440
Qy 4441 TAGATGTTATTTAAATAATGACGAATATCACTCTTATTGACAAATACCTAAATATGAG 4500
Db 4441 TAGATGTTATTTAAATAATGACGAATATCACTCTTATTGACAAATACCTAAATATGAG 4500
Qy 4501 TTTTATTAATTTAAGACTGTAATGTTCTTAAACCACTAACTACTGAGAGCTCAATG 4560
Db 4501 TTTTATTAATTTAAGACTGTAATGTTCTTAAACCACTAACTACTGAGAGCTCAATG 4560
Qy 4561 ATTGACATCTGAAATGCTTTGTAATTTATTTGACTTCAGGCCCTTAAGAACTGATTTCA 4620
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTATTTGACTTCAGGCCCTTAAGAACTGATTTCA 4620
Qy 4621 CGTGAGGCTTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAAGTGAATATGTTT 4680
Db 4621 CGTGAGGCTTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAAGTGAATATGTTT 4680

Qy 4681 TGGAAATGTRACATATTCAAAACAGAGTGCTCATTTTAGAAATGAGTAGTGTGATGGCA 4740
Db 4681 TGGAAATGTRACATATTCAAAACAGAGTGCTCATTTTAGAAATGAGTAGTGTGATGGCA 4740
Qy 4741 CTGGCAATTAACAGTGGTGTCTTTTAAATACATTCATTGGTATATTCAGTAGTATCTCT 4800
Db 4741 CTGGCAATTAACAGTGGTGTCTTTTAAATACATTCATTGGTATATTCAGTAGTATCTCT 4800
Qy 4801 CTGAGTTGGTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGGCTGGTAGT 4860
Db 4801 CTGAGTTGGTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGGCTGGTAGT 4860
Qy 4861 AAATATTGAGGCCACCTGTGTCTTTTGTATACATCTCTTCTGCTGTTGTTTGTAGTTGT 4920
Db 4861 AAATATTGAGGCCACCTGTGTCTTTTGTATACATCTCTTCTGCTGTTGTTTGTAGTTGT 4920
Qy 4921 TTTTTCCTCAAAACCCCTCTAAATATGTAACCAATGTTTACGCTTGCAGCTGTACAAA 4980
Db 4921 TTTTTCCTCAAAACCCCTCTAAATATGTAACCAATGTTTACGCTTGCAGCTGTACAAA 4980
Qy 4981 ACTGCCACACAGCAGATGTGACCTTCAGCCCATATTTGCCAATCAGTGAATATTTT 5040
Db 4981 ACTGCCACACAGCAGATGTGACCTTCAGCCCATATTTGCCAATCAGTGAATATTTT 5040
Qy 5041 TTGTTGTTGTTGTTGTTGTTTGTGAGACAGATCTCTCTGTTGCCAGGCTGGAG 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTTGTGAGACAGATCTCTCTGTTGCCAGGCTGGAG 5100
Qy 5101 TGAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTCTGTC 5160
Db 5101 TGAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTCTGTC 5160
Qy 5161 TCAGCTTCTGAGTAGTGGGACTACAGGTGCAATGCCACACACCTGCTTAATTTTGTGA 5220
Db 5161 TCAGCTTCTGAGTAGTGGGACTACAGGTGCAATGCCACACACCTGCTTAATTTTGTGA 5220
Qy 5221 TTTTATGAGACGGGGTTTCCACCATATTTGTCAGGCTTATCTTGAACCTCTGACCTC 5280
Db 5221 TTTTATGAGACGGGGTTTCCACCATATTTGTCAGGCTTATCTTGAACCTCTGACCTC 5280
Qy 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTACAGGATAGCCAGTGCAC 5340
Db 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTACAGGATAGCCAGTGCAC 5340
Qy 5341 CCAGCCGAGAAATTTATTTTATGATGTTTAAACCTTTGGGCTCTAGCCATATTTATG 5400
Db 5341 CCAGCCGAGAAATTTATTTTATGATGTTTAAACCTTTGGGCTCTAGCCATATTTATG 5400
Qy 5401 TCATAATACATGATTTGTGAAGACAGATTTCCATGAGTAACTCTGACAGTATTTAG 5460
Db 5401 TCATAATACATGATTTGTGAAGACAGATTTCCATGAGTAACTCTGACAGTATTTAG 5460
Qy 5461 ATCATGATCTCAACAATATTTCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520
Db 5461 ATCATGATCTCAACAATATTTCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520
Qy 5521 ATGTAAATACGTGTTGCTGTAGAGTGTGATATTTCAAACTGAAATCTCATATA 5580
Db 5521 ATGTAAATACGTGTTGCTGTAGAGTGTGATATTTCAAACTGAAATCTCATATA 5580
Qy 5581 AAGTTAAATTTTGAATA 5597
Db 5581 AAGTTAAATTTTGAATA 5597

RESULT 4
AB033034
LOCUS
DEFINITION Homo sapiens mRNA for KIAA1208 protein, partial cds.
ACCESSION AB033034
VERSION AB033034.2 GI:20521791
KEYWORDS

AB033034 4511 bp mRNA linear PRI 10-MAY-2002
Homo sapiens mRNA for KIAA1208 protein, partial cds.

Db 1201 CGCCAGTTGAGTCTCAATCTTGGATTTGCAACTGGAAATGGAGACATCACTTTGAA 1260
Qy 2342 AGGATACAAATTTGTCCTCAAGTCAGCTTGTCTGAGATCATTTCTGATGAATCAGAGCATGC 2401
Db 1261 AGGATACAAATTTGTCCTCAAGTCAGCTTGTCTGAGATCATTTCTGATGAATCAGAGCATGC 1320
Qy 2402 TAAATATAAATAAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACA 2461
Db 1321 TAAATATAAATAAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACA 1380
Qy 2462 GGAAACAGGTTCAATTAAGCATCTCCCAACAGCTTAGAGTGCTCTGAAAGATTGCA 2521
Db 1381 GGAAACAGGTTCAATTAAGCATCTCCCAACAGCTTAGAGTGCTCTGAAAGATTGCA 1440
Qy 2522 GAGGTTGACTTTTCTCGAGTGAGTGTAAGAGTGAATGGTCAATGACCAAGGTCAGATCC 2581
Db 1441 GAGGTTGACTTTTCTCGAGTGAGTGTAAGAGTGAATGGTCAATGACCAAGGTCAGATCC 1500
Qy 2582 ACCCTGACATTTGGAGACCAAGCAGAGATTTAGAGTGAAACTCACACCCCAAAACCAAT 2641
Db 1501 ACCCTGACATTTGGAGACCAAGCAGAGATTTAGAGTGAAACTCACACCCCAAAACCAAT 1560
Qy 2642 AGCGGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGAT 2701
Db 1561 AGCGGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGAT 1620
Qy 2702 GACAAAGAAAGAAATCAACAGGGAAGAAAGAGAAACAGTAGATGGAGGAAATGC 2761
Db 1621 GACAAAGAAAGAAATCAACAGGGAAGAAAGAGAAACAGTAGATGGAGGAAATGC 1680
Qy 2762 TGAAATACATAGCGGTTACTGAGTGTTACTTGGAAAGAGCTGACAGATTACACAGA 2821
Db 1681 TGAAATACATAGCGGTTACTGAGTGTTACTTGGAAAGAGCTGACAGATTACACAGA 1740
Qy 2822 TAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAATTTTCCAAGATCTTCTCGACGA 2881
Db 1741 TAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAATTTTCCAAGATCTTCTCGACGA 1800
Qy 2882 AGAAGAGTCATTGAAAGACAAATGGCATCTTCTGATGACCAAAATATCGGAGGCA 2941
Db 1801 AGAAGAGTCATTGAAAGACAAATGGCATCTTCTGATGACCAAAATATCGGAGGCA 1860
Qy 2942 ACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTT 3001
Db 1861 ACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTT 1920
Qy 3002 TGGATTCAATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGATGATGATGATGATGAT 3061
Db 1921 TGGATTCAATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGATGATGATGATGATGAT 1980
Qy 3062 GCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGCTCATTTTCAAAAGTGG 3121
Db 1981 GCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGCTCATTTTCAAAAGTGG 2040
Qy 3122 CCATCTGAGGATATGCAAGTTTGGCTTCTCTTATTTTATTTATTTATTTATTTATTTATTTAT 3181
Db 2041 CCATCTGAGGATATGCAAGTTTGGCTTCTCTTATTTTATTTATTTATTTATTTATTTATTTAT 2100
Qy 3182 GCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGCTCATTTTCAAAAGTGG 3241
Db 2101 GCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGCTCATTTTCAAAAGTGG 2160
Qy 3242 TGACAGAGAAATCCGAACTGGCTACCAAGATTTCAAGAACTGGCTTAAAGTTGACGGA 3301
Db 2161 TGACAGAGAAATCCGAACTGGCTACCAAGATTTCAAGAACTGGCTTAAAGTTGACGGA 2220
Qy 3302 TTTGACAGGTCGGAACACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCAC 3361
Db 2221 TTTGACAGGTCGGAACACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCAC 2280
Qy 3362 GCAGCTTAAATATTTTCCACCAACTCAGAAATCTTACTATGATCCCAACTGCGACCGGT 3421
Db 2281 GCAGCTTAAATATTTTCCACCAACTCAGAAATCTTACTATGATCCCAACTGCGACCGGT 2340

Qy 3422 CACTAAAGTCTAGTAACAAACTGTAACAGTAAGTGAACAAATCCCAAGCATATATA 3481
Db 2341 CACTAAAGTCTAGTAACAAACTGTAACAGTAAGTGAACAAATCCCAAGCATATATA 2400
Qy 3482 GGCACAAACAAATATAGGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGAT 3541
Db 2401 GGCACAAACAAATATAGGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGAT 2460
Qy 3542 TCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGAA 3601
Db 2461 TCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGAA 2520
Qy 3602 GTTGTGTTGCTGATGACAAACATTCACACACATCATAAAGATGCTCAGACAGTGAAGC 3661
Db 2521 GTTGTGTTGCTGATGACAAACATTCACACACATCATAAAGATGCTCAGACAGTGAAGC 3580
Qy 3662 TGTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAG 3721
Db 2581 TGTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAG 2540
Qy 3722 AGAGTATGCAAAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGAA 3781
Db 2641 AGAGTATGCAAAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGAA 2700
Qy 3782 CAATTTGAAAGTTTGGACCCATTTGCTACTAGCAACATTTGATTTATCTTACTATATCTC 3841
Db 2701 CAATTTGAAAGTTTGGACCCATTTGCTACTAGCAACATTTGATTTATCTTACTATATCTC 2760
Qy 3842 ATTTTGTCTGAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAAAGGAGGATACA 3901
Db 2761 ATTTTGTCTGAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAAAGGAGGATACA 2820
Qy 3902 CAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCATCTAC 3961
Db 2821 CAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCATCTAC 2880
Qy 3962 CTGAGCATTTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGCTTTTGTGATGTGATG 4021
Db 2881 CTGAGCATTTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGCTTTTGTGATGTGATG 2940
Qy 4022 CTTAGCAGTTTGGCCGGAAGAAATATCAGATACATGCTGTTTGTGCGCATGAAT 4081
Db 2941 CTTAGCAGTTTGGCCGGAAGAAATATCAGATACATGCTGTTTGTGCGCATGAAT 3000
Qy 4082 ATAGCCACCTGACTAGGAATTTTAAACCAACCCACTGAAACCTTGTGTGCGAGAGCT 4141
Db 3001 ATAGCCACCTGACTAGGAATTTTAAACCAACCCACTGAAACCTTGTGTGCGAGAGCT 3060
Qy 4142 CTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTTTTATAAAAAG 4201
Db 3061 CTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTTTTATAAAAAG 3120
Qy 4202 GCTCAGTACAGAGACAGCTGTTAATTTCCACAGCAATCATTTGAGAGTAACTTATTT 4261
Db 3121 GCTCAGTACAGAGACAGCTGTTAATTTCCACAGCAATCATTTGAGAGTAACTTATTT 3180
Qy 4262 AGGAGAGCCCTATGCCAGCTGGGAGTGAATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAAA 4321
Db 3181 AGGAGAGCCCTATGCCAGCTGGGAGTGAATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAAA 3240
Qy 4322 GCCTTTGTCTAAAGTTTGTGACATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4381
Db 3241 GCCTTTGTCTAAAGTTTGTGACATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3299
Qy 4382 GTTAACTAGTATTTCTTGTCTTCTGAGTATAACGAATTTGGGATGCTCTAAACCTATTTTAT 4441
Db 3300 GTTAACTAGTATTTCTTGTCTTCTGAGTATAACGAATTTGGGATGCTCTAAACCTATTTTAT 3359
Qy 4442 AGATGTTATTTAAATATGAGCAATATACCTCTTTATTTGCAATACCTTAAATTTATGAGT 4501
Db 3360 AGATGTTATTTAAATATGAGCAATATACCTCTTTATTTGCAATACCTTAAATTTATGAGT 3419

QY 572 GCCATGCTTGTACTGACCCAGCCCTGCCAGCAACATCACCTGTAGGACGTGCCATC 631
DB 541 GCCAATGCTGTCTCGACCCAGCCCTGCCAGCAACATCACCTGTAGGACGTGCCATC 600
QY 632 TCTTTATCTCTTCTTCTTCTGTCAGTGTGATTTTCAATGTTGCAAAACCAAAACCC 691
DB 601 TCTTTATCTCTTCTTCTTCTGTCAGTGTGATTTTCAATGTTGCAAAACCAAAACCC 660
QY 632 TTCTACCAATGCTCTGATTTGTTTGTGACAGTACTAAGGATGTTGAAGTGTGCCATC 751
DB 661 TTCTACCAATGCTCTGATTTGTTTGTGACAGTACTAAGGATGTTGAAGTGTGCCATC 720
QY 752 TGGAGTCTTAAAGAAATAGCAGACAGACAGTATGAGGGGGTACTTGACAACAGATAA 811
DB 721 TGGATCTGTTAAAGAAATAGCAGACAGACAGTATGAGGGGGTACTTGACAACAGATAA 780
QY 812 AGAAGTCCCTGATTTAGTGTGATTAATGCAAGATTTGGCTTTCTGAGTGTGATTTCAACCAAC 871
DB 781 AGAAGTCCCTGATTTAGTGTGATTAATGCAAGATTTGGCTTTCTGAGTGTGATTTCAACCAAC 840
QY 872 ATTCAAGAAACCAATCAACTAAACAAACAAATGTCAGAAATCTTTCTCTAAAGTCAA 931
DB 841 ATTCAAGAAACCAATCAACTAAACAAACAAATGTCAGAAATCTTTCTCTAAAGTCAA 900
QY 932 ACTGTTGAGTGTGATTTAGAGGCCAGTGTAGCGCTTCTTAAACTGAAATACCCCAAGGA 991
DB 901 ACTGTTGAGTGTGATTTAGAGGCCAGTGTAGCGCTTCTTAAACTGAAATACCCCAAGGA 960
QY 992 TTTTCAAGAAATGATTAAGCAACTAAGAAACATGACATGATGGAAGAAAGTAC 1051
DB 961 TTTTCAAGAAATGATTAAGCAACTAAGAAACATGACATGATGGAAGAAAGTAC 1020
QY 1052 CATAGTCTGTCATATTTATTTATGGAATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGA 1111
DB 1021 CATAGTCTGTCATATTTATTTATGGAATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGA 1080
QY 1112 AGACATCTGTCAGTCTGTTTGAAGATTAAGAAACATGAGTACTCATTTGCGATCTAT 1171
DB 1081 AGACATCTGTCAGTCTGTTTGAAGATTAAGAAACATGAGTACTCATTTGCGATCTAT 1140
QY 1172 CGAGAGGATGACCAATGGTTTCGGAATATTTTCAATGTCACCAACGGCGAGATTTCCATC 1231
DB 1141 CGAGAGGATGACCAATGGTTTCGGAATATTTTCAATGTCACCAACGGCGAGATTTCCATC 1200
QY 1232 CTGGCTGAACTCTGCAATCTCTGAGTGCATATAGTAACACACCAAGATGTTTTTCGAAA 1291
DB 1201 CTGGCTGAACTCTGCAATCTCTGAGTGCATATAGTAACACACCAAGATGTTTTTCGAAA 1260
QY 1292 TTTGAGCCACTGCTTACCTTTAGTTCACCTGCTATTGAAAGTGCATTTCAATCCATCGA 1351
DB 1261 TTTGAGCCACTGCTTACCTTTAGTTCACCTGCTATTGAAAGTGCATTTCAATCCATCGA 1320
QY 1352 AGGGCTGTCCAGAAAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCGT 1411
DB 1321 AGGGCTGTCCAGAAAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCGT 1380
QY 1412 GCCAGATGATTTTACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTGTGCC 1471
DB 1381 GCCAGATGATTTTACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTGTGCC 1440
QY 1472 AAACCTGTCCGAGGCTGCCAGGTTCTGTGATTAAGGATGGCTATTGTGCAAGGCTTG 1531
DB 1441 AAACCTGTCCGAGGCTGCCAGGTTCTGTGATTAAGGATGGCTATTGTGCAAGGCTTG 1500
QY 1532 TAAATATTCAGCTGCGATTTGGATTTGGTGGGATTTGCTCTGAAAACAGTGGAGGATCG 1591
DB 1501 TAAATATTCAGCTGCGATTTGGATTTGGTGGGATTTGCTCTGAAAACAGTGGAGGATCG 1560
QY 1592 CTATATTCAGAGGTTGGAGTACTGGGATTTGGAGTTGGACACCCCTGGCATTTGG 1651
DB 1561 CTATATTCAGAGGTTGGAGTACTGGGATTTGGAGTTGGACACCCCTGGCATTTGG 1620

QY 1652 TGGAGGAATAAACAGTGTCTCTTACTGTAAATCAGGATGTGGAATTCCTGGCTCGCTGA 1711
DB 1621 TGGAGGAATAAACAGTGTCTCTTACTGTAAATCAGGATGTGGAATTCCTGGCTCGCTGA 1680
QY 1712 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTCTGTGGTTTGTGCTGGGACCTGTGG 1771
DB 1681 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTCTGTGGTTTGTGCTGGGACCTGTGG 1740
QY 1772 GCAAGATCAATTTTTCATGAATTTGATAAAGTGTATCCCTTCCCAAAACACAGCTACATAT 1831
DB 1741 GCAAGATCAATTTTTCATGAATTTGATAAAGTGTATCCCTTCCCAAAACACAGCTACATAT 1800
QY 1832 TATTTCCAAAGGTGAATGCTGCTTATTTTCAGCTTTTGAGAGTGTAGCCAAAGAGGAGT 1891
DB 1801 TATTTCCAAAGGTGAATGCTGCTTATTTTCAGCTTTTGAGAGTGTAGCCAAAGAGGAGT 1860
QY 1892 TGAAGTGTCTTATGTCAGCAATCCCAATTTTGCAGATGCTTCTTATTTGCCAAACAGTGGAA 1951
DB 1861 TGAAGTGTCTTATGTCAGCAATCCCAATTTTGCAGATGCTTCTTATTTGCCAAACAGTGGAA 1920
QY 1952 AACCATCCCTCATTAATGCACAGTGAATGAATGCCACCAACATAATTTTAAATCTCAC 2011
DB 1921 AACCATCCCTCATTAATGCACAGTGAATGAATGCCACCAACATAATTTTAAATCTCAC 1980
QY 2012 GTTTCAAAATACAAAACATGAAGATTTCAAAATGCAGATAACAGTGGAGGTGCACACAAG 2071
DB 1981 GTTTCAAAATACAAAACATGAAGATTTCAAAATGCAGATAACAGTGGAGGTGCACACAAG 2040
QY 2072 GGAAGGACCAAACTGAAATTTCTACGGCCAGAGGGTTTACGAAAATTTAGTTAGTCCCAT 2131
DB 2041 GGAAGGACCAAACTGAAATTTCTACGGCCAGAGGGTTTACGAAAATTTAGTTAGTCCCAT 2100
QY 2132 AACACTTCTTCCAGAGCGGAAATCTTTTGTAGGATATTTCCAAAAGAAAACGCTTCCC 2191
DB 2101 AACACTTCTTCCAGAGCGGAAATCTTTTGTAGGATATTTCCAAAAGAAAACGCTTCCC 2160
QY 2192 GAAGTTTAAAGACATGATGTTAACTCAACAGAGGAGCCAGGAAAGAGTGAATTTCC 2251
DB 2161 GAAGTTTAAAGACATGATGTTTAACTCAACAGAGGAGCCAGGAAAGAGTGAATTTCC 2220
QY 2252 CCTGGTAAATATTTTCACTCTCTTCCAAAAGACGCTTGTGAGTCTCAATACCTTGGATT 2311
DB 2221 CCTGGTAAATATTTTCACTCTCTTCCAAAAGACGCTTGTGAGTCTCAATACCTTGGATT 2280
QY 2312 GCAACTGGAACTGGAGACATCACTTTGAAAGATACAATTTGTCCAAGTCAAGCTTGCT 2371
DB 2281 GCAACTGGAACTGGAGACATCACTTTGAAAGATACAATTTGTCCAAGTCAAGCTTGCT 2340
QY 2372 GAGATCATTTCTGATGAACCTCACAGCATGCTAAATATAAATAATCAAGCTATAATAACAGA 2431
DB 2341 GAGATCATTTCTGATGAACCTCACAGCATGCTAAATATAAATAATCAAGCTATAATAACAGA 2400
QY 2432 TGAACCAATGACAGTGTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATCTTGCC 2491
DB 2401 TGAACCAATGACAGTGTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATCTTGCC 2460
QY 2492 AAACAGCTTAGGAGTGTCTGAAAGATTTGCAGAGGTTTGACTTTTCTCGAGTGAAGTGA 2551
DB 2461 AAACAGCTTAGGAGTGTCTGAAAGATTTGCAGAGGTTTGACTTTTCTCGAGTGAAGTGA 2520
QY 2552 AGTGAATGGTTCATGACAGGGTCAAGATCCACCCCTGGACTTGGAGACCAACAGAT 2611
DB 2521 AGTGAATGGTTCATGACAGGGTCAAGATCCACCCCTGGACTTGGAGACCAACAGAT 2580
QY 2612 TAGAGTGGAACTCACACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAAGCCCC 2671
DB 2581 TAGAGTGGAACTCACACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAAGCCCC 2640
QY 2672 ATCTCTGATTTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGA 2731
DB 2641 ATCTCTGATTTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGA 2700
QY 2732 AAAAGAACACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGGCGTTACTGAAGTGT 2791

Db	2701	AAAAGAGAACAGTAGAATGGAGGAAAATGCTGAAAATACACATAGGCGTTACTGAAGTGT	2761
Qy	2792	ACTTGGAGAAAGCTGCAGCATTAACACAGATAGTTACTTGGGCTTTTGGCATGGGAGAA	2851
Db	2761	ACTTGGAGAAAGCTGCAGCATTAACACAGATAGTTACTTGGGCTTTTGGCATGGGAGAA	2820
Qy	2852	AAAAAGTATTTCCAAGATCTTCTCGAGGAAGAAGATCATTTGAGACACAAATTGGCAT	2911
Db	2821	AAAAAAGTATTTCCAAGATCTTCTCGAGGAAGAAGATCATTTGAGACACAAATTGGCAT	2880
Qy	2912	CTTCACTGATACGAAAAATACTGGAGGCAACTAAAAAGATACATTTGCAGATCCCTCAG	2971
Db	2881	CTTCACTGATACGAAAAATACTGGAGGCAACTAAAAAGATACATTTGCAGATCCCTCAG	2940
Qy	2972	ATATGTAATAAATTCCTAAATAGCAAGTTTGGATTACATTCGCGGAAAGTCCCTGCTCA	3031
Db	2941	ATATGTAATAAATTCCTAAATAGCAAGTTTGGATTACATTCGCGGAAAGTCCCTGCTCA	3000
Qy	3032	CATGCGCTCACATGATTCGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA	3091
Db	3001	CATGCGCTCACATGATTCGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA	3060
Qy	3092	ATTTGACAGAGCTCATTTCAAAAGTCGCGCATTTGAGGATATGAGTTCGCTTCTC	3151
Db	3061	ATTTGACAGAGCTCATTTCAAAAGTCGCGCATTTGAGGATATGAGTTCGCTTCTC	3120
Qy	3152	TTATTTTATTTATCTCATGAGTCAGTCGAGCCACTGCAATATATCTCAAGTCTTTGATGA	3211
Db	3121	TTATTTTATTTATCTCATGAGTCAGTCGAGCCACTGCAATATATCTCAAGTCTTTGATGA	3180
Qy	3212	AGTTGATACAGATCAATCTGGTGCTTGTCTGCAGAGAAATCCGAAACACTGGGTACCAG	3271
Db	3181	AGTTGATACAGATCAATCTGGTGCTTGTCTGCAGAGAAATCCGAAACACTGGGTACCAG	3240
Qy	3272	AAATTCAGAACTGCGGTAAAGTTTTCAGGATTTTGACAGGCTCGGAACACATGCTAATAA	3331
Db	3241	AAATTCAGAACTGCGGTAAAGTTTTCAGGATTTTGACAGGCTCGGAACACATGCTAATAA	3300
Qy	3332	TTGCTCAAAAATGCTTCTGCTGATATCACGAGCTAAATAATTTCCACCAACTCAGGA	3391
Db	3301	TTGCTCAAAAATGCTTCTGCTGATATCACGAGCTAAATAATTTCCACCAACTCAGGA	3360
Qy	3392	ATCCTACTATGATCCCAACCTGCCACCGGTCACTAABAGTCTAGTAACAACTGTAAC	3451
Db	3361	ATCCTACTATGATCCCAACCTGCCACCGGTCACTAABAGTCTAGTAACAACTGTAAC	3420
Qy	3452	AGTAACATGACAAAAATCCACAAGCATATAAGGACAAAAAATAATATAGGTTTGGAAATCAT	3511
Db	3421	AGTAACATGACAAAAATCCACAAGCATATAAGGACAAAAAATAATATAGGTTTGGAAATCAT	3480
Qy	3512	GGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAGCTTCTCATGCTGTGGCCCA	3571
Db	3481	GGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAGCTTCTCATGCTGTGGCCCA	3540
Qy	3572	GTTGGATGACATAGAAAAAACCCCTAGGAAGTTTGTTCGCTGAAATGACAACTTGACCA	3631
Db	3541	GTTGGATGACATAGAAAAAACCCCTAGGAAGTTTGTTCGCTGAAATGACAACTTGACCA	3600
Qy	3632	CAATCATAAAGATGCTCAGACAGTGAAGGTGTTCTCAGGACCTTCTATGAATCCATGTT	3691
Db	3601	CAATCATAAAGATGCTCAGACAGTGAAGGTGTTCTCAGGACCTTCTATGAATCCATGTT	3660
Qy	3692	CCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAAAACCGTTTCCTTCATATGCA	3751
Db	3661	CCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAAAACCGTTTCCTTCATATGCA	3720
Qy	3752	TGAGCTCAGGATGGAGGCTTATCAGACAAATTTGAAGTTTGGACCACTTGTGACT	3811
Db	3721	TGAGCTCAGGAAATGGAGGCTTATCAGACAAATTTGAAGTTTGGACCACTTGTGACT	3780
Qy	3812	AGCAACATTCGATATGTTTACTATATCTCATTTTTTTCGTGAGCAGTTAATTCACCTTAA	3871

Db	3781	AGCAACATTCATATGTTTACTATATCTCATTTTCTTGCTGAGCAGTAAATTCGACTTAA	384
Qy	3872	CGCGAAGATATTTCCCGAAGAGGAGTACACAAGAAGCTAGTCCCAATCGAATCAGAGT	3931
Db	3841	CGCGAAGATATTTCCCGAAGAGGAGTACACAAGAAGCTAGTCCCAATCGAATCAGAGT	3900
Qy	3932	ATAGAAGATCTTCATTTGAAACCCATCTCCTCAGCATTTTACTGAGCATTTTAAAACTCA	3991
Db	3901	ATAGAAGATCTTCATTTGAAACCCATCTCAGCATTTTACTGAGCATTTTAAAACTCA	3960
Qy	3992	GCITTCAGAGATGCTCTTTGTGATGTATGCTGTAGCAGTTTGGCCCGAAGAGGAAAAATA	4051
Db	3961	GCITTCAGAGATGCTCTTTGTGATGTATGCTGTAGCAGTTTGGCCCGAAGAGGAAAAATA	4020
Qy	4052	TCCAGTACCATGCTGTTTGTGGCATGAATATAGCCCATCTGACTAGGAATATTATTAACCA	4111
Db	4021	TCCAGTACCATGCTGTTTGTGGCATGAATATAGCCCATCTGACTAGGAATATTATTAACCA	4080
Qy	4112	ACCCATGAAAACCTGTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGAATTGCG	4171
Db	4081	ACCCATGAAAACCTGTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGAATTGCG	4140
Qy	4172	TCATGGACCTGTGCATCCTTTTATAAAAAGGCTCAGTCAAGAGACAGCTGTTAATTTC	4231
Db	4141	TCATGGACCTGTGCATCCTTTTATAAAAAGGCTCAGTCAAGAGACAGCTGTTAATTTC	4200
Qy	4232	CCACAGCATCATTCAGACTAACTTTATAGAGAGGCTATGCCAGCTGGGAGTGATT	4291
Db	4201	CCACAGCATCATTCAGACTAACTTTATAGAGAGGCTATGCCAGCTGGGAGTGATT	4260
Qy	4292	GCTAAGAGGCTCCAGTCTTTGCATCCAAAGCCTTTTGCTAAAGTTTGCATTTT-TT	4350
Db	4261	GCTAAGAGGCTCCAGTCTTTGCATCCAAAGCCTTTTGCTAAAGTTTGCATTTTCT	4320
Qy	4351	TTTTTCATTTCCCATTTTAAAGTAGTTA-CTAAGTTAACTAGTTATT	4395
Db	4321	TTTTTCATTTCCCATTTTCAAGTAGTTACTTAAGTTAACTAGTTACT	4366
RESULT 6			
AX468106		4277 bp	DNA
LOCUS			linear
DEFINITION	Sequence 29 from Patent WO0250279.		PAT 16-JUL-2002
ACCESSION	AX468106		
VERSION	AX468106.1	GI:21900980	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Baughin, M.R., Lu, Y., Arvizu, C., Ramkumar, J., Yao, M.G., Policky, J.L., Wallia, N.K., Tribouley, K.M., Yue, H., Batra, S., Ding, J., Lal, P.G., Borowsky, M.L., Lu, D.A., Gandhi, A.R., Griffin, J.A., Xu, X., Azmzal, Y., Gietzen, K.J., Tang, Y.F., Warren, B.A., Mason, P.M., Burford, N., Hafalia, A.J., Lee, E.A., Yang, J., Gorvad, A.E., Emerling, B.M., Marquis, J.P., Lee, S.Y., Swarnakar, A. and Reddy, R.		
TITLE	Nucleic acid-associated proteins		
JOURNAL	Patent: WO 0250279-A 29 27-JUN-2002;		
FEATURES	Incyte Genomics, Inc. (US)		
source	Location/Qualifiers		
	1..4277		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN	/notes="Incyte ID No: 7503560CB1"		

ORIGIN

Figure 6

10

1

QY 4406 AGTATAACGAATTGGGATGCTCTAAB-CCTATTTTATAGATGTTATTTAAATAATGTCAGC 4464
Db 4202 AGTATAACGAATTGGGATGCTCTAABCCCTATTTTATAGATGTTATTTAAATAATGTCAGC 4261
QY 4465 AATATCACCTCTATT 4480
Db 4262 ATTTNCACCTCTTTT 4277

RESULT 7

LOCUS AR300471 3783 bp DNA linear PAT 12-JUN-2003
DEFINITION Sequence 20 from patent US 6537785.
ACCESSION AR300471
VERSION AR300471.1 GI:31687913
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

Unclassified.

1 (bases 1 to 3783)

AUTHORS Canfield, W.M.

TITLE Methods of treating lysosomal storage diseases

JOURNAL Patent: US 6537785-A 20 25-MAR-2003;

FEATURES Location/Qualifiers

source 1..3783

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match 67.2%; Score 3761; DB 6; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 165 ATGCTGTTCAAGCTCCTGCGAGACAAACCTATACCTGCTGCCAGGATGGGCTC 224
Db 7 ATGGGGTTCAGCTCTTGCGAGACAAACCTATACCTGCTGCCAGGATGGGCTC 66
QY 225 TACGTGTGCTTTGGGGCTGCTGTCCATCGTCTCGCTGCCAGTTCGGAGAGTG 284
Db 67 TACGTGTGCTTTGGGGCTGCTGTCCATCGTCTCGCTGCCAGTTCGGAGAGTG 126
QY 285 GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAAT 344
Db 127 GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAAT 186
QY 345 GCTGGAAGTCTTTTCAAGATCGGCTTTGTCTGCCATGCCGATGAGCTGTTTACACC 404
Db 187 GCTGGAAGTCTTTTCAAGATCGGCTTTGTCTGCCATGCCGATGAGCTGTTTACACC 246
QY 405 TGGGTGAATGGCAGATCTTGAATCTAGAGGAATCTACAGAGGTACAGAGACAGATG 464
Db 247 TGGGTGAATGGCAGATCTTGAATCTAGAGGAATCTACAGAGGTACAGAGACAGATG 306
QY 465 GAGGAGGACAGAAAGCAATGAGAAATCCCTTGGGAAAAACACAAACGAAACCTACTAAG 524
Db 307 GAGGAGGACAGAAAGCAATGAGAAATCCCTTGGGAAAAACACAAACGAAACCTACTAAG 366
QY 525 AAGAGTGAGACAGTGTAGTGTGTTGCTTAACACTGCTAAGGTGCCAATGCTTGTGA 584
Db 367 AAGAGTGAGACAGTGTAGTGTGTTGCTTAACACTGCTAAGGTGCCAATGCTTGTGC 426
QY 585 CTGGACCCAGCCCTGCGAGCAACATCACCTTGAAGGACGTGCCATCTCTTATCCTTCT 644
Db 427 CTGGACCCAGCCCTGCGAGCAACATCACCTTGAAGGACGTGCCATCTCTTATCCTTCT 486
QY 645 TTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704
Db 487 TTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546
QY 705 TCAGTGTGTTTGTGACAGTACTAAGGATGTTGAAGTGCCTGCTGCTGCTGCTTAA 764
Db 547 TCAGTGTGTTTGTGACAGTACTAAGGATGTTGAAGTGCCTGCTGCTGCTTAA 606

QY 765 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGATATAAGAGTCCCTGGA 824
Db 607 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGTATAAGAGTCCCTGGA 666
QY 825 TTAGTGTAAATGCAAGATTTGGCTTTCTAGTGGATTTCCACCAACATTTCAAGGAAACA 884
Db 667 TTAGTGTAAATGCAAGATTTGGCTTTCTAGTGGATTTCCACCAACATTTCAAGGAAACA 726
QY 885 AATCACTAAACAAATGCGAGAAATCTTCTCTAAAGTCAAACTGTTGCGAGTTG 944
Db 727 AATCACTAAACAAATGCGAGAAATCTTCTCTAAAGTCAAACTGTTGCGAGTTG 786
QY 945 TATTGAGAGCCAGTGTAGCGCTTTTAAACCTGAAATAACCCCAAGGATTTTCAAGAAATTG 1004
Db 787 TATTGAGAGCCAGTGTAGCGCTTTTAAACCTGAAATAACCCCAAGGATTTTCAAGAAATTG 846
QY 1005 AATAAGCAAACTAAGAAACATGACATGATGAAAGAACTGACCATTAAGTCTCTGCA 1064
Db 847 AATAAGCAAACTAAGAAACATGACATGATGAAAGAACTGACCATTAAGTCTCTGCA 906
QY 1065 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124
Db 907 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966
QY 1125 AGTCGTTTGAAGATAACGAAAGAACTGAGGTACTCATTTGCGATCTATTCGAGAGGATGCA 1184
Db 967 AGTCGTTTGAAGATAACGAAAGAACTGAGGTACTCATTTGCGATCTATTCGAGAGGATGCA 1026
QY 1185 CCATGGTTCGGAATATTTTCAATGTCACCAACGGCAGATTCATCTCGCTGGCTGAACCTT 1244
Db 1027 CCATGGTTCGGAATATTTTCAATGTCACCAACGGCAGATTCATCTCGCTGGCTGAACCTT 1086
QY 1245 GACAACTCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1304
Db 1087 GACAACTCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1146
QY 1305 CCTACCTTTAGTTTCACTGCTATTTGAAGTCAATTCATCGCATCAAGGGCTGTCCAG 1364
Db 1147 CCTACCTTTAGTTTCACTGCTATTTGAAGTCAATTCATCGCATCAAGGGCTGTCCAG 1206
QY 1365 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1424
Db 1207 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1266
QY 1425 TACAGTCACTCCAAAGCCAGAAAGGTTTATTTGACATGCGCTGTGCCAACTGTGCCGAG 1484
Db 1267 TACAGTCACTCCAAAGCCAGAAAGGTTTATTTGACATGCGCTGTGCCAACTGTGCCGAG 1326
QY 1485 GGTGCCCCAGGTTTCCCTGGATTAAAGGATGGCTATTTGTGACAAAGGCTTGTAAATTCAGCC 1544
Db 1327 GGTGCCCCAGGTTTCCCTGGATTAAAGGATGGCTATTTGTGACAAAGGCTTGTAAATTCAGCC 1386
QY 1545 TGCATTTGGATCGTGGGATTCCTCTGGAACACAGTGGAGGGAGTCCGTATATTCAGGGA 1604
Db 1387 TGCATTTGGATCGTGGGATTCCTCTGGAACACAGTGGAGGGAGTCCGTATATTCAGGGA 1446
QY 1605 GGTGGAGTACTCGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAAC 1664
Db 1447 GGTGGAGTACTCGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAAC 1506
QY 1665 AGTGTCTTCTTACTGTAATCAGGATGCGAAATTCCTGGCTCGCTGATAGTTTCTGTGAC 1724
Db 1507 AGTGTCTTCTTACTGTAATCAGGATGCGAAATTCCTGGCTCGCTGATAGTTTCTGTGAC 1566
QY 1725 CAAGCATGAATGTCCTTGTCTGTGGTTGATGCTGGGACCTGTGGGCAAGATCAATTT 1784
Db 1567 CAAGCATGAATGTCCTTGTCTGTGGTTGATGCTGGGACCTGTGGGCAAGATCAATTT 1626
QY 1785 CATGAATTTGTAAGAGTCTTCTCCAAACACAGACTCACTATATTTTCCAAAAGGT 1844
Db 1627 CATGAATTTGTAAGAGTCTTCTTCCAAACACAGACTCACTATATTTTCCAAAAGGT 1686
QY 1845 GAATGCTGCTTTATTTACGCTTTTGAGAGTAGCCAAAGAGAGGATTTGAAGGTGCTAT 1904

Db 1687 GAATCCCTGCGCTATTTTTCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1746
Qy 1905 AGTGACAAATCCAAATATTCGACATGCTTCTATTGCGCAACAAAGTGGAAACCATCCACCTC 1964
Db 1747 AGTGACAAATCCAAATATTCGACATGCTTCTATTGCGCAACAAAGTGGAAACCATCCACCTC 1806
Qy 1965 ATATGCGACAGTGGAAATGCGCACCAATACATTTTAACTCTCAGCTTTCAAAATACA 2024
Db 1807 ATAAATGCGACAGTGGAAATGCGCACCAATACATTTTAACTCTCAGCTTTCAAAATACA 1866
Qy 2025 AACGATGAAGAGTTCAAAATGCGACATAACAGTGGAGTGGACACAAAGGGAGGACCAAAA 2084
Db 1867 AACGATGAAGAGTTCAAAATGCGACATAACAGTGGAGTGGACACAAAGGGAGGACCAAAA 1926
Qy 2085 CTGAATCTCAGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCAATAACACTTCTTCCA 2144
Db 1927 CTGAATCTCAGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCAATAACACTTCTTCCA 1986
Qy 2145 GAGCGGAAATCCTTTTGGAGATATCCCAAGAAAGAAACGCTTCCCGAAGTTTAAAGAGA 2204
Db 1987 GAGCGGAAATCCTTTTGGAGATATCCCAAGAAAGAAACGCTTCCCGAAGTTTAAAGAGA 2046
Qy 2205 CATGATGTTAACTCAACAGGAGAGCCCGAGGAAGAGTGAAATTTCCCTGGTAAATATT 2264
Db 2047 CATGATGTTAACTCAACAGGAGAGCCCGAGGAAGAGTGAAATTTCCCTGGTAAATATT 2106
Qy 2265 TCACCTCTCCAAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCACTGGACAT 2324
Db 2107 TCACCTCTCCAAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCACTGGACAT 2166
Qy 2325 GGAGACATCACTTTGAAAGGATACAAATTTGTCAGAGTCCAGCTTGGATCATATTCTG 2384
Db 2167 GGAGACATCACTTTGAAAGGATACAAATTTGTCAGAGTCCAGCTTGGATCATATTCTG 2226
Qy 2385 ATGAATCTCAGAGCATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAATGAC 2444
Db 2227 ATGAATCTCAGAGCATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAATGAC 2286
Qy 2445 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAATAAAGCATCTTGCCAAACAGCTTAGGA 2504
Db 2287 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAATAAAGCATCTTGCCAAACAGCTTAGGA 2346
Qy 2505 GTGCTGAAAGATTCAGAGGTTGACTTTTCTCGAGTGAAGTGAAGTGAATGGTCAAT 2564
Db 2347 GTGCTGAAAGATTCAGAGGTTGACTTTTCTCGAGTGAAGTGAAGTGAATGGTCAAT 2406
Qy 2565 GACAGGGTCAAGATCCACCCCTGAGCTTGAGACACACAGCAAGATTTAGAGTGGAACT 2624
Db 2407 GACAGGGTCAAGATCCACCCCTGAGCTTGAGACACACAGCAAGATTTAGAGTGGAACT 2466
Qy 2625 CACACCCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCCCATCTCTGATTTGTT 2684
Db 2467 CACACCCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCCCATCTCTGATTTGTT 2526
Qy 2685 CCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAAAGAACAGT 2744
Db 2527 CCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAAAGAACAGT 2586
Qy 2745 AGAATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAAGTGTACTTGGAAAGAAAG 2804
Db 2587 AGAATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAAGTGTACTTGGAAAGAAAG 2646
Qy 2805 CTGACGATTAACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTC 2864
Db 2647 CTGACGATTAACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTC 2706
Qy 2865 CAAATCTCTCGAGGAGAGAGTCAATTTGAAGACAAATTTGGCATATCTTCACTGTATGC 2924
Db 2707 CTGATCTCTCGAGGAGAGAGTCAATTTGAAGACAAATTTGGCATATCTTCACTGTATGC 2766
Qy 2925 AAAAATCTGGAGGCAACTAAAGATACATTTGCGATTTCCCTCAGATATGTAATAAA 2984

Db 2767 AAGNATCTGGAGGCACTAAAGATACATTTGCGAGATCCCTCAGATATGTAATAAA 2826
Qy 2985 ATTCTAAATAGCAAGTTTGGATTCACTCGCGGAAAGTCCCTGCTCACTGCTCACAATG 3044
Db 2827 ATTCTAAATAGCAAGTTTGGATTCACTCGCGGAAAGTCCCTGCTCACAATGCTCACAATG 2886
Qy 3045 ATTGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGCAAGACG 3104
Db 2887 ATTGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGCAAGACG 2946
Qy 3105 TCATTTCAAAAGTGGCCATTCTGAGGATATGCAAGTTTGGCTTCTCTTTATTTATTTAT 3164
Db 2947 TCATTTCAAAAGTGGCCATTCTGAGGATATGCAAGTTTGGCTTCTCTTTATTTATTTAT 3006
Qy 3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224
Db 3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066
Qy 3225 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCTACCAGAAATTCAGAACTG 3284
Db 3067 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCTACCAGAAATTCAGAACTG 3126
Qy 3285 CCGTTAAGTTTGCAGGATTTGACAGTCTGGAACACATGCTTAATAAATGCTCAAAATG 3344
Db 3127 CCGTTAAGTTTGCAGGATTTGACAGTCTGGAACACATGCTTAATAAATGCTCAAAATG 3186
Qy 3345 CTTCTGCTGATATCACGAGCTAAATAATATTCCACCAACTCAGGAATCTCTACTATGAT 3404
Db 3187 CTTCTGCTGATATCACGAGCTAAATAATATTCCACCAACTCAGGAATCTCTACTATGAT 3246
Qy 3405 CCCAACCTGCCACCCTGCTCACTAAAGTCTAGTAAACAACTGTAAACCACTGTAACCTGACAAA 3464
Db 3247 CCCAACCTGCCACCCTGCTCACTAAAGTCTAGTAAACAACTGTAAACCACTGTAACCTGACAAA 3306
Qy 3465 ATCCACAGAGCATATAAGGACAAACAAATATATGTTTGAATCATCGGAGAGAAAGAA 3524
Db 3307 ATCCACAGAGCATATAAGGACAAACAAATATATGTTTGAATCATCGGAGAGAAAGAA 3366
Qy 3525 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATA 3584
Db 3367 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATA 3426
Qy 3585 AGAAAAACCTTAGAAGTTTGGTTCCTGATGATGACCAACTGACCAACTATTAAGAT 3644
Db 3427 AGAAAAACCTTAGAAGTTTGGTTCCTGATGATGACCAACTGACCAACTATTAAGAT 3486
Qy 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3704
Db 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3546
Qy 3705 CAAATTTGAACCTCCAAAGAGATTCGAAAACGTTTCCCTCATATGTCATGAGCTGACAGAA 3764
Db 3547 CAAATTTGAACCTCCAAAGAGATTCGAAAACGTTTCCCTCATATGTCATGAGCTGACAGAA 3606
Qy 3765 TCGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACTTGAAT 3824
Db 3607 TCGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACTTGAAT 3666
Qy 3825 ATGTTTACTATATCTCATTTTCTGCTGAGCAGTTTAAATGACCTTAAACGGAAGATATTT 3884
Db 3667 ATGTTTACTATATCTCATTTTCTGCTGAGCAGTTTAAATGACCTTAAACGGAAGATATTT 3726
Qy 3885 CCCAGAGGAGATACACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAAGATC 3941
Db 3727 CCCAGAGGAGATACACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAAGATC 3783

RESULT 8
AR428748
LOCUS
DEFINITION Sequence 20 from patent US 6642038.
ACCESSION AR428748

AR428748 3783 bp DNA linear PAT 18-DEC-2003

KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 3783)
AUTHORS Canfield, W.M.
TITLE GlcNAc phosphotransferase of the lysosomal targeting pathway
JOURNAL Patent: US 6642039-A 20 04-NOV-2003;
FEATURES Location/Qualifiers
source 1..3783
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 67.2%; Score 3761; DB 6; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 165 ATGCTGTTCAAGCTCCTCGAGAGACAAACCTATACCTCGCTGTCCACAGGTATGGGCTC 224
DB 7 ATGGGTTTCAGCTCTTCGAGAGACAAACCTATACCTCGCTGTCCACAGGTATGGGCTC 66
QY 225 TACGTGTGCTTCTGGGGTGGTGTGTCACCAATGCTGCTCCGCTTCCAGTTCGGAGAGGTG 284
DB 67 TACGTGTGCTTCTGGGGTGGTGTGTCACCAATGCTGCTCCGCTTCCAGTTCGGAGAGGTG 126
QY 285 GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATT 344
DB 127 GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATT 186
QY 345 GCTGAAAGTCCCTTTACAGATCGGCTTTGTCTGCCCATGCCGATGAGCTGTTTACACC 404
DB 187 GCTGAAAGTCCCTTTACAGATCGGCTTTGTCTGCCCATGCCGATGAGCTGTTTACACC 246
QY 405 TGGTGAATGGACAGATCTTGAATCTAGAGNACTACAGCAGGTACAGAGACAGATG 464
DB 247 TGGTGAATGGACAGATCTTGAATCTAGAGNACTACAGCAGGTACAGAGACAGATG 306
QY 465 GAGGAGGAGCGAAGCAATGAGAGAAATCCTTGGGAAAAACACAAACGACCTTACTAAG 524
DB 307 GAGGAGGAGCGAAGCAATGAGAGAAATCCTTGGGAAAAACACAAACGACCTTACTAAG 366
QY 525 AAGAGTGAAGCAGTGTAGAGTGTGTGTAAACACTGCATTAAGTGCCAAATGCTGTGA 584
DB 367 AAGAGTGAAGCAGTGTAGAGTGTGTGTAAACACTGCATTAAGTGCCAAATGCTGTGC 426
QY 585 CTGGACCCAGCCCTGCCAGCAACATCACCTGAGGACGTGCCATCTCTTTATCCTTCT 644
DB 427 CTGGACCCAGCCCTGCCAGCAACATCACCTGAGGACCTGCCATCTCTTTATCCTTCT 486
QY 645 TTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704
DB 487 TTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546
QY 705 TCAGTGTGTTTTTGACAGTACTAAGGATGTTGAGATGCCACTCTGGACTGCTTAAA 764
DB 547 TCAGTGTGTTTTTGACAGTACTAAGGATGTTGAGATGCCACTCTGGACTGCTTAAA 606
QY 765 GGAATAGCAGACACAGATGAGAGGGGTACTTGACACAGATAAAGAGTCCCTGGA 824
DB 607 GGAATAGCAGACACAGATGAGAGGGGTACTTGACACAGATAAAGAGTCCCTGGA 666
QY 825 TTAGTGCTAAAGCAAGATTTGGCTTCTCTGAGTGAATTTCCACCAACATTCAGGAAACA 884
DB 667 TTAGTGCTAAAGCAAGATTTGGCTTCTCTGAGTGAATTTCCACCAACATTCAGGAAACA 726
QY 885 AATCAACTAAACCAAAATTCGCAAAAATCTTCTCTAAAGTCAAACTGTCAGTTG 944
DB 727 AATCAACTAAACCAAAATTCGCAAAAATCTTCTCTAAAGTCAAACTGTCAGTTG 786
QY 945 TATTCAGAGGCGAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 1004
DB 787 TATTCAGAGGCGAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 846

QY 1005 AATAAGCAAACTAAGAAACATGACCATGTATGTAAGAAAAAGAACTGACCATAGTCTCTGCA 1064
DB 847 AATAAGCAAACTAAGAAACATGACCATGTATGTAAGAAAAAGAACTGACCATAGTCTCTGCA 906
QY 1065 TATTATTATTGGGATCTGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124
DB 907 TATTATTATTGGGATCTGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966
QY 1125 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTCGGATCTATTCAGAGGCGCATGCA 1184
DB 967 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTCGGATCTATTCAGAGGCGCATGCA 1026
QY 1185 CCATGGTTCGGAAATATTTTTCATGTCACCAAGGCGCATTCATCTCTGGTGAACCTT 1244
DB 1027 CCATGGTTCGGAAATATTTTTCATGTCACCAAGGCGCATTCATCTCTGGTGAACCTT 1086
QY 1245 GACAATCCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1304
DB 1087 GACAATCCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1146
QY 1305 CCTACCTTTAGTTCACTGCTATTTGAAAGTCACTCATCGCATCGAAGGCGTGTCCGAG 1364
DB 1147 CCTACCTTTAGTTCACTGCTATTTGAAAGTCACTCATCGCATCGAAGGCGTGTCCGAG 1206
QY 1365 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGATGTCCTGGCCACAGATGTTTT 1424
DB 1207 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGATGTCCTGGCCACAGATGTTTT 1266
QY 1425 TACAGTCACTCCAAAAGGCGAGAGTTTATTTGACATGCGCTGTGCAAACTGTGCGGAG 1484
DB 1267 TACAGTCACTCCAAAAGGCGAGAGTTTATTTGACATGCGCTGTGCAAACTGTGCGGAG 1326
QY 1485 GGTGCCCCAGGTTCTCGGATTAAGGATGCTATTGTAACAAGCTGTGTAATTAATTCAGCC 1544
DB 1327 GGTGCCCCAGGTTCTCGGATTAAGGATGCTATTGTAACAAGCTGTGTAATTAATTCAGCC 1386
QY 1545 TGCATTTGGGATGGTGGGATTTGCTCTGAAAACAGTGGAGGAGTGCCTATATTGCAAGA 1604
DB 1387 TGCATTTGGGATGGTGGGATTTGCTCTGAAAACAGTGGAGGAGTGCCTATATTGCAAGA 1446
QY 1605 GGTGAGGATCTCGGAGTATTTGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAAC 1664
DB 1447 GGTGAGGATCTCGGAGTATTTGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAAC 1506
QY 1665 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1724
DB 1507 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1566
QY 1725 CAAGCATGCAATGTCCTTGTCTGTGGGTTTGTATGCTGGCGACTGTGGGCAAGATCATTTT 1784
DB 1567 CAAGCATGCAATGTCCTTGTCTGTGGGTTTGTATGCTGGCGACTGTGGGCAAGATCATTTT 1626
QY 1785 CATGAATTTGTATAAAGTGTATCCTTCTCCCAAAACAGACTCACTATATATTATCCAAAGGT 1844
DB 1627 CATGAATTTGTATAAAGTGTATCCTTCTCCCAAAACAGACTCACTATATATTATCCAAAGGT 1686
QY 1845 GAATGCTGCTTTTATTTCAGCTTTTGCAAGATGAGCCAAAGAGGAGTTGAAGGTGCTAT 1904
DB 1687 GAATGCTGCTTTTATTTCAGCTTTTGCAAGATGAGCCAAAGAGGAGTTGAAGGTGCTAT 1746
QY 1905 AGTGACAATCCATATTTGCAATGCTTCTATTTGCCAAACAGTGGAAACCATCCACCTC 1964
DB 1747 AGTGACAATCCATATTTGCAATGCTTCTATTTGCCAAACAGTGGAAACCATCCACCTC 1806
QY 1965 ATAATGCACAGTGGAAATGAATGCCACCAATACATTTTAACTCTCAGTTTCAAAATACA 2024
DB 1807 ATAATGCACAGTGGAAATGAATGCCACCAATACATTTTAACTCTCAGTTTCAAAATACA 1866
QY 2025 AACGNTGAGAGTTTCAAAATGCAAGATAACAGTGGAGTGGACACAAAGGAGGACCAAAA 2084
DB 1867 AACGNTGAGAGTTTCAAAATGCAAGATAACAGTGGAGTGGACACAAAGGAGGACCAAAA 1926

Best Local Similarity 99.7%; Pred. No. 0;				
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;				
QY	165	ATGCTGTTCAAGCTCCTCGCAGACAAACCTATACCTCGCTGTCCTCCACAGGTATGGGCTC	224	
Db	7	ATGGGTTCAAGCTCTTGCAGAGACAAACCTATACCTCGCTGTCCTCCACAGGTATGGGCTC	66	
QY	225	TACGTGTCTCTTGGGGTGGTGTGTACACCTGCTCCGCTTCCAGTTCGGAGAGGTG	284	
Db	67	TACGTGTCTCTTGGGGTGGTGTGTACACCTGCTCCGCTTCCAGTTCGGAGAGGTG	126	
QY	285	GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATT	344	
Db	127	GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATT	186	
QY	345	GCTGAAAAGTCTTTCAAGATCGGCTTGTCTGCCATGCGGATGACGTTGTTTACACC	404	
Db	187	GCTGAAAAGTCTTTCAAGATCGGCTTGTCTGCCATGCGGATGACGTTGTTTACACC	245	
QY	405	TGGGTGAATGGACAGATCTTGAACCTAAGAGAACTACAGAGGTGAGAGACAGATG	464	
Db	247	TGGGTGAATGGACAGATCTTGAACCTAAGAGAACTACAGAGGTGAGAGACAGATG	306	
QY	465	GAGGAGGAGCAGAAAGCAATGAGAGAAATCTTGGGAAAAACACACGGAACCTTACTAAG	524	
Db	307	GAGGAGGAGCAGAAAGCAATGAGAGAAATCTTGGGAAAAACACACGGAACCTTACTAAG	366	
QY	525	AAGAGTGAAAGCAGTATAGAGTGTGTTGTAACACACTGCATTAAGGTGCCAATGCTGTA	584	
Db	367	AAGAGTGAAAGCAGTATAGAGTGTGTTGTAACACACTGCATTAAGGTGCCAATGCTGTC	426	
QY	585	CTGGACCCAGCCCTGCCAGCCAACTACCTCGAGGAGCGTGCCTCTCTTTATCCCTCT	644	
Db	427	CTGGACCCAGCCCTGCCAGCCAACTACCTCGAGGAGCGTGCCTCTCTTTATCCCTCT	486	
QY	645	TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTACCAATGTC	704	
Db	487	TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTACCAATGTC	546	
QY	705	TCAGTGTGTTTTCAGAGTACTAAGATGTTGAAGATGTCCTGCTGCTGCTGCTGCTGCT	764	
Db	547	TCAGTGTGTTTTCAGAGTACTAAGATGTTGAAGATGTCCTGCTGCTGCTGCTGCTGCT	606	
QY	765	GGAAATAGCAGACACAGTATGGAGGGGTACTTGACACAGATAAAGAAAGTCCCTGGA	824	
Db	607	GGAAATAGCAGACACAGTATGGAGGGGTACTTGACACAGATAAAGAAAGTCCCTGGA	666	
QY	825	TTAGTGCTAATGCAAGATTTGGCTTTCCTGAGTGGATTCACCAACATTCAGAGAAACA	884	
Db	667	TTAGTGCTAATGCAAGATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGAGAAACA	726	
QY	885	AATCAACTAAAACAAAATTCGCAAAATCTTCTCTTAAGTCAAACTGCTGCTGCTG	944	
Db	727	AATCAACTAAAACAAAATTCGCAAAATCTTCTCTTAAGTCAAACTGCTGCTGCTG	786	
QY	945	TATTCAGAGGCCAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATG	1004	
Db	787	TATTCAGAGGCCAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATG	846	
QY	1005	AATAAGCAAACTAAGAGAAACATGACATTTGATGAAAGAACTGACCAATAGTCTGCA	1064	
Db	847	AATAAGCAAACTAAGAGAAACATGACATTTGATGAAAGAACTGACCAATAGTCTGCA	906	
QY	1065	TATTTATTATGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC	1124	
Db	907	TATTTATTATGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC	966	
QY	1125	AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGTGA	1184	
Db	967	AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGTGA	1026	
QY	1185	CCATGGGTTCGAATATTTTCAATGTCACCAAGGGAGATTCATCTCTGGCTGAACCTT	1244	

Db	1027	CCATGGGTTTCGGAATATTTTTCATTTGTACCAAGGGCAGATTCCTCTGGCTGAACCTT	1086	
QY	1245	GACAACTCCGAGTGAACAATAGTAACACACACAGGATGTTTTTTCGAAATTTGACCACTTG	1304	
Db	1087	GACAACTCCGAGTGAACAATAGTAACACACACAGGATGTTTTTTCGAAATTTGACCACTTG	1146	
QY	1305	CCTACCTTTAGTTCACTGCTATTGAAAGTCAATTCAGCATCGCATCGAAGGCTGTCCCAG	1364	
Db	1147	CCTACCTTTAGTTCACTGCTATTGAAAGTCAATTCAGCATCGCATCGAAGGCTGTCCCAG	1206	
QY	1365	AAGTTTATTACCTAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT	1424	
Db	1207	AAGTTTATTACCTAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT	1266	
QY	1425	TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCCGAG	1484	
Db	1267	TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCCGAG	1326	
QY	1485	GGCTGCCAGGTTCTCGGATTAAGGATGCTATTGTAACAAGCTTGTATAATTAATCAGCC	1544	
Db	1327	GGCTGCCAGGTTCTCGGATTAAGGATGCTATTGTAACAAGCTTGTATAATTAATCAGCC	1386	
QY	1545	TGCGATTGGGATGGTGGGGATTGCTCTGGAAAACAGTGGAGGGAGTCCCTATATTGCGAG	1604	
Db	1387	TGCGATTGGGATGGTGGGGATTGCTCTGGAAAACAGTGGAGGGAGTCCCTATATTGCGAG	1446	
QY	1605	GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC	1664	
Db	1447	GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC	1506	
QY	1665	AGTGTCTCTTACTGTAATCAGGGATGTCGAAATTCCTGGCTGCTCATTAAGTTCTGTGAC	1724	
Db	1507	AGTGTCTCTTACTGTAATCAGGGATGTCGAAATTCCTGGCTGCTCATTAAGTTCTGTGAC	1566	
QY	1725	CAAGCATGCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1784	
Db	1567	CAAGCATGCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1626	
QY	1785	CATGAATGCTATAAAGTGATCTCTTCCCAACACAGACTCACTATATTATTCCTCAAGGT	1844	
Db	1627	CATGAATGCTATAAAGTGATCTCTTCCCAACACAGACTCACTATATTATTCCTCAAGGT	1686	
QY	1845	GAATGCTGCTTATTTTTCAGCTTTGCGAAGTAGCCAAAAGAGGAGTTGAAGGTGCTAT	1904	
Db	1687	GAATGCTGCTTATTTTTCAGCTTTGCGAAGTAGCCAAAAGAGGAGTTGAAGGTGCTAT	1746	
QY	1905	AGTGACATCCAAATTAATGACATGCTTCTATTGCGCAACAGTGGAAAAACCATCCACCTC	1964	
Db	1747	AGTGACATCCAAATTAATGACATGCTTCTATTGCGCAACAGTGGAAAAACCATCCACCTC	1806	
QY	1965	ATAATGCACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTTTCAAAAATACA	2024	
Db	1807	ATAATGCACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTTTCAAAAATACA	1866	
QY	2025	AACGATGAAGAGTTCAAAATGACAGATAACAGTGGAGGTGGACACAAGAGGAGGACAAA	2084	
Db	1867	AACGATGAAGAGTTCAAAATGACAGATAACAGTGGAGGTGGACACAAGAGGAGGACAAA	1926	
QY	2085	CTGAATCTACGGCCAGAGGGTTACGAAATTTTAGTTAGTCCCATTAACATCTTCTCCA	2144	
Db	1927	CTGAATCTACGGCCAGAGGGTTACGAAATTTTAGTTAGTCCCATTAACATCTTCTCCA	1986	
QY	2145	GAGCGGAAATCTTTTGGAGATATTCCTCAAGAAAAACGCTTCCCGAAGTTTAAAGAGA	2204	
Db	1987	GAGCGGAAATCTTTTGGAGATATTCCTCAAGAAAAACGCTTCCCGAAGTTTAAAGAGA	2046	
QY	2205	CATGATGTTAACTCAACAGGAGAGCGGAGGAGGTGAANAATTCCTGCTGTAATATT	2264	
Db	2047	CATGATGTTAACTCAACAGGAGAGCGGAGGAGGTGAANAATTCCTGCTGTAATATT	2106	
QY	2265	TCACCTCTTCCAAAAGACGCCAGTGTGAGTCTCAATACCTTTGGATTGCAACTGGAACAT	2324	
Db	2107	TCACCTCTTCCAAAAGACGCCAGTGTGAGTCTCAATACCTTTGGATTGCAACTGGAACAT	2166	

QY 2325 GGAGACATCACTTTGAAGGATACAAATTTGTCCAAAGTCAGCGCTTGCTGAGATCATTTCTG 2384
DB |||||
QY 2167 GGAGACATCACTTTGAAGGATACAAATTTGTCCAAAGTCAGCGCTTGCTGAGATCATTTCTG 2226
DB |||||
QY 2385 ATGAACCTCAGAGCATGTAAATTAATAAATCAAGCTATAATAACAGATGAAGAAACAATGAC 2444
DB |||||
QY 2227 ATGAACCTCAGAGCATGTAAATTAATAAATCAAGCTATAATAACAGATGAAGAAACAATGAC 2286
DB |||||
QY 2445 AGTTTGGTGGCTCCACAGGAAACAGCTTCATATAAAGACATCTTGCCAAACAGCTTAGGA 2504
DB |||||
QY 2287 AGTTTGGTGGCTCCACAGGAAACAGCTTCATATAAAGACATCTTGCCAAACAGCTTAGGA 2346
DB |||||
QY 2505 GTGTCTGAAGATTGACAGAGTTGACCTTTCTGCACTGAGTGTAAGAAGTGAATGCTCAT 2564
DB |||||
QY 2347 GTGTCTGAAGATTGACAGAGTTGACCTTTCTGCACTGAGTGTAAGAAGTGAATGCTCAT 2406
DB |||||
QY 2555 GACCAAGGTCAAGATCCACCCCTGACCTTGAGACACACAGAGATTTAGAGTGGAACT 2624
DB |||||
QY 2407 GACCAAGGTCAAGATCCACCCCTGACCTTGAGACACACAGAGATTTAGAGTGGAACT 2466
DB |||||
QY 2625 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGACCCCATCTCTGATTGTT 2684
DB |||||
QY 2467 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGACCCCATCTCTGATTGTT 2526
DB |||||
QY 2685 CACTGGAAGCCAGATGACAAAAGAAAGAAATCAACAGGAAAGAAAGAAAGAAAGACAGT 2744
DB |||||
QY 2527 CACTGGAAGCCAGATGACAAAAGAAAGAAATCAACAGGAAAGAAAGAAAGAAAGACAGT 2586
DB |||||
QY 2745 AGAATGAGGAAATGCTGAAATACATAGGCGTCTTACTGAAGTGTACTTGAAGAAAG 2804
DB |||||
QY 2587 AGAATGAGGAAATGCTGAAATACATAGGCGTCTTACTGAAGTGTACTTGAAGAAAG 2646
DB |||||
QY 2805 CTGACGATTAACAGATAGTTACTTGGGCTTTTGGCGATGCGGAGAAAGAAATGATTTTC 2864
DB |||||
QY 2647 CTGACGATTAACAGATAGTTACTTGGGCTTTTGGCGATGCGGAGAAAGAAAGAAATGATTTTC 2706
DB |||||
QY 2865 CAGATCTCTCGACGAGAGAGTCATTTGAAGACACAAATTTGGCATACTTCACCTGATAGC 2924
DB |||||
QY 2707 CTAGATCTTCTGACGAAAGAGATGCTTTGAAGACACAAATTTGGCATACTTCACCTGATAGC 2766
DB |||||
QY 2925 AAAAATCTGGAGGCACTAAAAGATACATTTCCAGATATCCCTCAGATATGTAATAA 2984
DB |||||
QY 2767 AAGAACTCTGGAGGCACTAAAAGATACATTTCCAGATATCCCTCAGATATGTAATAA 2826
DB |||||
QY 2985 ATCTTAATAGCAAGTTTGGATTCACATCGGGAAGTCCCTGCTCACATGCTCCATG 3044
DB |||||
QY 2827 ATCTTAATAGCAAGTTTGGATTCACATCGGGAAGTCCCTGCTCACATGCTCCATG 2886
DB |||||
QY 3045 ATTGACCGATTGTTATCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 3104
DB |||||
QY 2887 ATTGACCGATTGTTATCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 2946
DB |||||
QY 3105 TCATTTCAAAAGTGGCCATCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3164
DB |||||
QY 2947 TCATTTCAAAAGTGGCCATCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3006
DB |||||
QY 3165 CTCATGAGTGCAGTGCAGCCACTGAATATCTCAAGTCTTTTATGATGATTTGATACAGAT 3224
DB |||||
QY 3007 CTCATGAGTGCAGTGCAGCCACTGAATATCTCAAGTCTTTTATGATGATTTGATACAGAT 3066
DB |||||
QY 3225 CAATCTGTGTCTGTGACAGAGAAATCCGAACCTGGCTACAGAAATTCAGAACTG 3284
DB |||||
QY 3067 CAATCTGTGTCTGTGACAGAGAAATCCGAACCTGGCTACAGAAATTCAGAACTG 3126
DB |||||
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAATTCCTCAAAATG 3344
DB |||||
QY 3127 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAATTCCTCAAAATG 3186
DB |||||
QY 3345 CTTCTGTGATATCAGCAGCTAAATAATTTCCACCACTCAGGAATCTTACTATGAT 3404
DB |||||
QY 3187 CTTCTGTGATATCAGCAGCTAAATAATTTCCACCACTCAGGAATCTTACTATGAT 3246
DB |||||

QY 3405 CCCAACCTGCCACCGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAACCTGACAAA 3464
DB |||||
QY 3247 CCCAACCTGCCACCGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAACCTGACAAA 3306
DB |||||
QY 3465 ATCCCAAGCATATAGGACAAACAAATATATAGTTTGAATCATGCGGAAGAGAA 3524
DB |||||
QY 3307 ATCCCAAGCATATAGGACAAACAAATATATAGTTTGAATCATGCGGAAGAGAA 3366
DB |||||
QY 3525 ATCGCTTTTAAATGATTTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3584
DB |||||
QY 3367 ATCGCTTTTAAATGATTTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3426
DB |||||
QY 3585 AGAAAAACCTTAGGAAGTTTGTTCCTGCTGAATGACACATTCAGCAATCATAAAGAT 3644
DB |||||
QY 3427 AGAAAAACCTTAGGAAGTTTGTTCCTGCTGAATGACACATTCAGCAATCATAAAGAT 3486
DB |||||
QY 3645 GCTCAGACAGTGAAGGCTTCTCAGGACATTCATGAATCCATGTTCCCATACCTTCC 3704
DB |||||
QY 3487 GCTCAGACAGTGAAGGCTTCTCAGGACATTCATGAATCCATGTTCCCATACCTTCC 3546
DB |||||
QY 3705 CAATTTGAATCTGCCAAGAGATATCGAAACCGTTTCTTCATATGCTAGCTGAGCTGAGAA 3764
DB |||||
QY 3547 CAATTTGAATCTGCCAAGAGATATCGAAACCGTTTCTTCATATGCTAGCTGAGCTGAGAA 3606
DB |||||
QY 3765 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3824
DB |||||
QY 3607 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3666
DB |||||
QY 3825 ATGTTTACTATATCTCAATTTTGTCTGAGCAGTTAATGCACTTAAGCGGAAGATATT 3884
DB |||||
QY 3667 ATGTTTACTATATCTCAATTTTGTCTGAGCAGTTAATGCACTTAAGCGGAAGATATT 3726
DB |||||
QY 3885 CCCAGAAGGAGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC 3941
DB |||||
QY 3727 CCCAGAAGGAGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC 3783
DB |||||

RESULT 10

AR300472 3621 bp DNA linear PAT 12-JUN-2003
LOCUS AR300472 Sequence 21 from patent US 6537785.

DEFINITION AR300472

ACCESSION AR300472

VERSION AR300472.1 GI:31687914

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

AUTHORS 1 (bases 1 to 3621)

TITLE Canfield, W.M.

JOURNAL Methods of treating lysosomal storage diseases

Patent: US 6537785-A 21 25-MAR-2003;

FEATURES Location/Qualifiers

Source

1..3621

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match 62.3%; Score 3485.4; DB 6; Length 3621;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATTGCTGGAAGTCC 356

DB |||||

QY 115 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATTGCTGGAAGTCC 174

DB |||||

QY 357 TTTCAGAAATCGGCTTTGTCTGCCATGCGGATTTGACGTTGTTTACACCTGGGTGAATGGC 416

DB |||||

QY 175 TTTCAGAAATCGGCTTTGTCTGCCATGCGGATTTGACGTTGTTTACACCTGGGTGAATGGC 234

DB |||||

QY 417 ACAGATCTTTGAACTACTGAAAGAACTACAGCAGGTACAGAAACAGATGCGAGAGAGCAG 476

DB |||||

QY 235 ACAGATCTTTGAACTACTGAAAGAACTACAGCAGGTACAGAAACAGATGCGAGAGAGCAG 294

DB |||||

QY 477 AAAGCAATGAGAGAAATCCTTGGGAAAAACAAACGGAACCTTACTAAGAGAGTGAGAG 536
DB 295 AAAGCAATGAGAGAAATCCTTGGGAAAAACAAACGGAACCTTACTAAGAGAGTGAGAG 354
QY 537 CAGTTAGAGTGTTCCTAACACACTGCATTAAAGTGGCCCAATGCTTGTACTGGACCCAGCC 596
DB 355 CAGTTAGAGTGTTCCTAACACACTGCATTAAAGTGGCCCAATGCTTGTACTGGACCCAGCC 414
QY 597 CTGCCAGCAACATCACCTGAGGAGCGTGCATCTCTTTATCTCTTTTCATTTCTGCTGCC 656
DB 415 CTGCCAGCAACATCACCTGAGGAGCGTGCATCTCTTTATCTCTTTTCATTTCTGCTGCC 474
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 716
DB 475 AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 534
QY 717 TTTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTTAAAGGAAATAGCAGA 776
DB 535 TTTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTTAAAGGAAATAGCAGA 594
QY 777 CAGACAGTATGAGGGGGTACTTTGACACAGATAAAGAAAGTCCCTGGATTAGTGCTAATG 836
DB 595 CAGACAGTATGAGGGGGTACTTTGACACAGATAAAGAAAGTCCCTGGATTAGTGCTAATG 654
QY 837 CAAAGATTTGGCTTTCTCTGAGTGGATTTCCACAAATTTCAAGGAAACAAATCAACTAAAA 896
DB 655 CAAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAATTTCAAGGAAACAAATCAACTAAAA 714
QY 897 ACABAATTTGCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTGCTGAGGCGCC 956
DB 715 ACABAATTTGCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTGCTGAGGCGCC 774
QY 957 AGTGTAGCGCTTCTAAAACTGAATTAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 1016
DB 775 AGTGTAGCGCTTCTAAAACTGAATTAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 834
QY 1017 AAGAGACATGACCAATGATGGAAGAAAGAACTGACCAATAGTCTGCAATATTATTATGG 1076
DB 835 AAGAGAAATGACCAATGATGGAAGAAAGAACTGACCAATAGTCTGCAATATTATTATGG 894
QY 1077 GATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 1136
DB 895 GATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 954
QY 1137 GATAACGAAGAACTGAGTACTCATTTGGGATCTATCGAGAGCGATGCAACCTTACCTCGA 1196
DB 955 GATAACGAAGAACTGAGTACTCATTTGGGATCTATCGAGAGCGATGCAACCTTACCTCGA 1014
QY 1197 AATATTTTCAATGTGCACCAACGGGAGATTTCCATCTCGCTGCAACCTTTGACAACTCTCGA 1256
DB 1015 AATATTTTCAATGTGCACCAACGGGAGATTTCCATCTCGCTGCAACCTTTGACAACTCTCGA 1074
QY 1257 GTGACAAATAGTAACACACAGAGATGTTTTTGGAAATTTGAGCCACTTCCCTACCTTTAGT 1316
DB 1075 GTGACAAATAGTAACACACAGAGATGTTTTTGGAAATTTGAGCCACTTCCCTACCTTTAGT 1134
QY 1317 TCACCTGCTATTGAAAGTACATTTTCATGCAATCGAAGGGCTGTCCGAGAGTTTATTATAC 1376
DB 1135 TCACCTGCTATTGAAAGTACATTTTCATGCAATCGAAGGGCTGTCCGAGAGTTTATTATAC 1194
QY 1377 CTAATATGATGATGATGATGTTTGGGAAAGATGTTCTGGCCAGATGATTTTACAGTCACTTCC 1436
DB 1195 CTAATATGATGATGATGATGTTTGGGAAAGATGTTCTGGCCAGATGATTTTACAGTCACTTCC 1254
QY 1437 AAAGCCGAAGAGTTTATTTGACATGGCTGTGCAAACTGTGCGAGGGGTGCGCCAGGT 1496
DB 1255 AAAGCCGAAGAGTTTATTTGACATGGCTGTGCAAACTGTGCGAGGGGTGCGCCAGGT 1314
QY 1497 TCCTGGATTAAGGATGGCTATTGTGACAGGGCTTGTATAATTTCAAGCTGCGATTGGGAT 1556
DB 1315 TCCTGGATTAAGGATGGCTATTGTGACAGGGCTTGTATAATTTCAAGCTGCGATTGGGAT 1374
QY 1557 GGTGGGGATGCTCTGGAACACAGTGGAGGAGTGCCTATATTGACAGGAGTGGAGGTACT 1616

DB 1375 GGTGGGGATGCTCTGGAACACAGTGGAGGAGTGCCTATATTGACAGGAGTGGAGGTACT 1434
QY 1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1676
DB 1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1494
QY 1677 TGTAAATCAGAGATGTGGAATTCCTGGCTCGCTGATAAGTTCTGTGCACCAAGCATGCAAT 1736
DB 1495 TGTAAATCAGAGATGTGGAATTCCTGGCTCGCTGATAAGTTCTGTGCACCAAGCATGCAAT 1554
QY 1737 GTCTGTGCTGTGGGTTTGATGTGCGAGTGTGGCAAGATCAATTTTCAATGAATTTGAT 1796
DB 1555 GTCTGTGCTGTGGGTTTGATGTGCGAGTGTGGCAAGATCAATTTTCAATGAATTTGAT 1614
QY 1797 AAGTGTATCTCTCCCAACACAGATCTCACTATATTATTTCCAAAAGGTGAATGCCCTGCC 1856
DB 1615 AAGTGTATCTCTCCCAACACAGATCTCACTATATTATTTCCAAAAGGTGAATGCCCTGCC 1674
QY 1857 TATTTTCAGCTTTTCAGAAAGTAGCAAAAGAGGAGTTGAAGTGCCTATAGTACAAATCCA 1916
DB 1675 TATTTTCAGCTTTTCAGAAAGTAGCAAAAGAGGAGTTGAAGTGCCTATAGTACAAATCCA 1734
QY 1917 ATAATTCGACATGCTTCTATTTGCCAACAAGTGGAATAAACAATCCACCTCATATATGACAGT 1976
DB 1735 ATAATTCGACATGCTTCTATTTGCCAACAAGTGGAATAAACAATCCACCTCATATATGACAGT 1794
QY 1977 GGAATGAATGCCACCAACATACATTTTATCTCACGTTTCAAAATACAAACGATGAAGAG 2036
DB 1795 GGAATGAATGCCACCAACATACATTTTATCTCACGTTTCAAAATACAAACGATGAAGAG 1854
QY 2037 TCCAAATCGAGATAACAGTGGAGTGGACACAAGGAGGAGCCAAACACTGAATTTCTACG 2096
DB 1855 TCCAAATCGAGATAACAGTGGAGTGGACACAAGGAGGAGCCAAACACTGAATTTCTACG 1914
QY 2097 GCCCAGAGGTTTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 2156
DB 1915 GCCCAGAGGTTTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 1974
QY 2157 CTTTTTGAAGATATTCCCAAGAAAAACGCTTCCCGAAGTTTAAAGAGACATGATGTTAAC 2216
DB 1975 CTTTTTGAAGATATTCCCAAGAAAAACGCTTCCCGAAGTTTAAAGAGACATGATGTTAAC 2034
QY 2217 TCAACAGGAGAGCCAGAGNAGAGGTGAATAATTTCCCTCGTGAATAATTTTCACTCTCTTCCA 2276
DB 2035 TCAACAGGAGAGCCAGAGNAGAGGTGAATAATTTCCCTCGTGAATAATTTTCACTCTCTTCCA 2094
QY 2277 AAAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTCGAACCTGGAAACATGGAGACATCACT 2336
DB 2095 AAAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTCGAACCTGGAAACATGGAGACATCACT 2154
QY 2337 TTGAAAGGATPACAAATTTGTCCAAAGTCAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAG 2396
DB 2155 TTGAAAGGATPACAAATTTGTCCAAAGTCAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAG 2214
QY 2397 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAATAACCAATGCAAGTTTGGTGGCT 2456
DB 2215 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAATAACCAATGCAAGTTTGGTGGCT 2274
QY 2457 CCACAGAAAAACAGGTTTCATAAAAGCATCTTGCCTGAAACAGCTTAGGAGTGTCTGAAAGA 2516
DB 2275 CCACAGAAAAACAGGTTTCATAAAAGCATCTTGCCTGAAACAGCTTAGGAGTGTCTGAAAGA 2334
QY 2517 TTCCAGAGTTGACTTTTCTCGAGTGTGTAAGATGAATGGTGTATGACCAAGGTCAG 2576
DB 2335 TTCCAGAGTTGACTTTTCTCGAGTGTGTAAGATGAATGGTGTATGACCAAGGTCAG 2394
QY 2577 AATCCACCCCTGGAGTGGAGACCAAGCAAGATTTTAGAGTGGAAACTCACACCCAAAAA 2636
DB 2395 AATCCACCCCTGGAGTGGAGACCAAGCAAGATTTTAGAGTGGAAACTCACACCCAAAAA 2454
QY 2637 ACCATAGGCGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTTCCATCTGGAAGC 2696

2455	ACCATAGCGCGAATGTGACAAAGAAAAAGCCCCCATCTCTGTGATTGTTTCACATGGAAAGC	2514
2497	CAGATGACAAAGAAAAAATAATCACAGGGAAGAAAAAGAGAACAGTAGAATGGAGGAA	2756
2515	CAGATGACAAAGAAAAAATAATCACAGGGAAGAAAAAGAGAACAGTAGAATGGAGGAA	2574
2757	AATCCTGAAATACATAGGCGTTACTTGAAGTGTACTTGGAGNAAGCTTGCAGCATTC	2816
2575	AATCCTGAAATACATAGGCGTTACTTGAAGTGTACTTGGAGNAAGCTTGCAGCATTC	2634
2817	ACAGATAGTTACTTTGGGCTTTTGGCCATGGAGAAAAAAAGTAATTTCCAAAGATCTTCTC	2876
2635	ACAGATAGTTACTTTGGGCTTTTGGCCATGGAGAAAAAAAGTAATTTCTTAGATCTTCTC	2694
2877	GACGAAGAAGAGTCATTGAAGACACAATTTGGCATACTTCTAGATGACAAAAATCTGGG	2936
2695	GACGAAGAAGAGTCATTGAAGACACAATTTGGCATACTTCTAGATGACAAAAATCTGGG	2754
2937	AGGCAACTAAAGATACATTTGCGAGTTCCCTCAGATATGTAAATAAAATTTCTAAATAGC	2996
2755	AGGCAACTAAAGATACATTTGCGAGTTCCCTCAGATATGTAAATAAAATTTCTAAATAGC	2814
2997	AAGTTTGGATTCACATCGCGAAAAGTCCCTGCTCACATGCCTCACATGATGACCGGATT	3056
2815	AAGTTTGGATTCACATCGCGAAAAGTCCCTGCTCACATGCCTCACATGATGACCGGATT	2874
3057	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGACGTCAATTCACAAA	3116
2875	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGACGTCAATTCACAAA	2934
3117	GTGCGCCATCTCGAGGATATGCAGTTTGGCCTTCTCTTAATTTTATTATCTCATGAGTGCA	3176
2935	GTGCGCCATCTCGAGGATATGCAGTTTGGCCTTCTCTTAATTTTATTATCTCATGAGTGCA	2994
3177	GTGCAGCCACTGAATATATCTCAAGTCCTTTGATGAAGTTGANTCAGATCAATCTGTGTCTC	3236
2995	GTGCAGCCACTGAATATATCTCAAGTCCTTTGATGAAGTTGANTCAGATCAATCTGTGTCTC	3054
3237	TTGTCGCACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCGTTAAGTTTG	3296
3055	TTGTCGCACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCGTTAAGTTTG	3114
3297	CAGGATTTGACAGGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT	3356
3115	CAGGATTTGACAGGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT	3174
3357	ATCAGCAGCTAAATAATATATCCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3416
3175	ATCAGCAGCTAAATAATATATCCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3234
3417	CCGCTGCATAAAGTCTAGTAACAAAACCTGTAAACCCAGTAACTGACAAAAATCCCAAGCA	3476
3235	CCGCTGCATAAAGTCTAGTAACAAAACCTGTAAACCCAGTAACTGACAAAAATCCCAAGCA	3294
3477	TATAAGGACAAAAACAATATAGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAA	3536
3295	TATAAGGACAAAAACAATATAGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAA	3354
3537	ATGATTGCTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAAACCTT	3596
3355	ATGATTGCTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAAACCTT	3414
3597	AGGAAGTTTGTTCGCTGAATGACAAATATGACCAAAATCATAAAGATGCTCAGACAGTG	3656
3415	AGGAAGTTTGTTCGCTGAATGACAAATATGACCAAAATCATAAAGATGCTCAGACAGTG	3474
3657	AAGCTGTCTCAGGACTTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGAACTG	3716
3475	AAGCTGTCTCAGGACTTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGAACTG	3534
3717	CCAAGAGAGTATCGAAACCGTTTCTCCTTCATATGATGAGCTGCAGGAATCGAGGGCTTAT	3776
3535	CTAGAGAGTATCGAAACCGTTTCTCCTTCATATGATGAGCTGCAGGAATCGAGGGCTTAT	3594

Qy	3777	CGAGACAAATTTGAAGTTTGGAC	3799
Db	3595	CGAGACAAATTTGAAGTAGTAGTC	3617
RESULT 11			
LOCUS	AR428749	3621 bp	DNA linear PAT 18-DEC-2003
DEFINITION	Sequence 21 from patent US 6642038.		
ACCESSION	AR428749		
VERSION	AR428749.1	GI:40198483	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 3621)		
TITLE	Canfield, W.M.		
JOURNAL	GlcNac phosphotransferase of the lysosomal targeting pathway		
FEATURES	Patent: US 6642038-A 21 04-NOV-2003;		
source	Location/Qualifiers		
	1..3621		
	/organism="unknown"		
ORIGIN	/mol_type="genomic DNA"		
	Query Match 62.3%; Score 3485.4; DB 6; Length 3621;		
	Best Local Similarity 99.7%; Pred. No. 0;		
	Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;		
Qy	297	AGCGAGATCAATACCACTGTTTGTGATTCCTATAGACACAATATTCGTGGAAAGTCC	356
Db	115	AGCGAGATCAATACCACTGTTTGTGATTCCTATAGACACAATATTCGTGGAAAGTCC	174
Qy	357	TTTCAGAAATCGGCTTTGCTGCCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGC	416
Db	175	TTTCAGAAATCGGCTTTGCTGCCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGC	234
Qy	417	ACAGATCTTGAACACTAGTGAAGAACTACAGAGGTGACAGAACAGATGGAGGAGGACAG	475
Db	235	ACAGATCTTGAACACTAGTGAAGAACTACAGAGGTGACAGAACAGATGGAGGAGGACAG	294
Qy	477	AAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGAGTGAGAAG	536
Db	295	AAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGAGTGAGAAG	354
Qy	537	CAGTTAGAGTGTTCATAACACACTGCATTAAGGTGCCAATGCTGTGACGCCAGCC	596
Db	355	CAGTTAGAGTGTTCATAACACACTGCATTAAGGTGCCAATGCTGTGACGCCAGCC	414
Qy	597	CTGCCACCAACATCACCCCTGAAGGACGTGCCCATCTCTTTATCCTTCTTTTCATTTCTGCC	655
Db	415	CTGCCACCAACATCACCCCTGAAGGACGTGCCCATCTCTTTATCCTTCTTTTCATTTCTGCC	474
Qy	657	AGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTGTTT	716
Db	475	AGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTGTTT	534
Qy	717	TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGGACTGCTTAAAGGAAATAGCAGA	776
Db	535	TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGGACTGCTTAAAGGAAATAGCAGA	594
Qy	777	CAGACAGTATGGAGGGGTACTTGCACAAACAGATAAAGAGTTCCTCGATAGTGTCTAATG	836
Db	595	CAGACAGTATGGAGGGGTACTTGCACAAACAGATAAAGAGTTCCTCGATAGTGTCTAATG	654
Qy	837	CAGATTTTGCTTTCTGAGTGGATTTCACCAACATTCGAAGGAAACAAATCAACTAAAA	896
Db	655	CAGATTTTGCTTTCTGAGTGGATTTCACCAACATTCGAAGGAAACAAATCAACTAAAA	714
Qy	897	ACAAATTTGCCAGAAAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGTTATTCAGAGGCC	956
Db	715	ACAAATTTGCCAGAAAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGTTATTCAGAGGCC	774

QY 957 AGTGTAGCCCTCTTAAACTGATATAACCCCAAGGATTTTCAAGATTAAGCAAACT 1016
DB 775 AGTGTAGCCCTCTTAAACTGATATAACCCCAAGGATTTTCAAGATTAAGCAAACT 834
QY 1017 AAGAAAGACATGACCAATGATGAAAGAACTGACCAATGACCAATGATGAAAGTATTAATG 1076
DB 835 AAGAAAGACATGACCAATGATGAAAGAACTGACCAATGATGAAAGTATTAATG 894
QY 1077 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCGCAGTCGTTTGAA 1136
DB 895 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCGCAGTCGTTTGAA 954
QY 1137 GATAACGAAGAACTGAGTACTCATTTGGATCTATCGAGAGCATGCAACCATGGTTCGG 1196
DB 955 GATAACGAAGAACTGAGTACTCATTTGGATCTATCGAGAGCATGCAACCATGGTTCGG 1014
QY 1197 AATATTTTCAATGTCACCAACGGGAGATTCATCCTCGGTGAACTTGAATTTATTTAC 1256
DB 1015 AATATTTTCAATGTCACCAACGGGAGATTCATCCTCGGTGAACTTGAATTTATTTAC 1074
QY 1257 GTGACAAATAGTAACACACAGGATGTTTTCGAAATTTGAGCAATTTGAGTACCTTTAGT 1316
DB 1075 GTGACAAATAGTAACACACAGGATGTTTTCGAAATTTGAGCAATTTGAGTACCTTTAGT 1134
QY 1317 TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTAC 1376
DB 1135 TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTAC 1194
QY 1377 CTAATATGATGTCATGTTGGGAAGGATGTCGCGCAGATGATTTTACAGTCACTCC 1436
DB 1195 CTAATATGATGTCATGTTGGGAAGGATGTCGCGCAGATGATTTTACAGTCACTCC 1254
QY 1437 AAGGCCAGAAAGTTTATTTGACATGGCTGTGCGCAACTGTGCGAGGGTGCACAGT 1496
DB 1255 AAGGCCAGAAAGTTTATTTGACATGGCTGTGCGCAACTGTGCGAGGGTGCACAGT 1314
QY 1497 TCCTGGAATTAAGATGGCTATTGTGCAAGGCTGTGTAATAATTCAGCCTGCGATGGGAT 1556
DB 1315 TCCTGGAATTAAGATGGCTATTGTGCAAGGCTGTGTAATAATTCAGCCTGCGATGGGAT 1374
QY 1557 GGTGGGATGCTCTGGAACAGTGGAGGAGTGCCTATATTGCAAGGAGTGGAGTACT 1616
DB 1375 GGTGGGATGCTCTGGAACAGTGGAGGAGTGCCTATATTGCAAGGAGTGGAGTACT 1434
QY 1617 GGGATATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1676
DB 1435 GGGATATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1494
QY 1677 TGTAATCAGGATGTCGAAATTCCTGGCTCGCTGATTAAGTTCCTGTGACCAAGATGCAT 1736
DB 1495 TGTAATCAGGATGTCGAAATTCCTGGCTCGCTGATTAAGTTCCTGTGACCAAGATGCAT 1554
QY 1737 GTCTTGTCTGTGGGTTTGTGCTGGGACTGTGGGCAAGATCATTTTCATGAATTTGTAT 1796
DB 1555 GTCTTGTCTGTGGGTTTGTGCTGGGACTGTGGGCAAGATCATTTTCATGAATTTGTAT 1614
QY 1797 AAGTGTACTTCTCCCAACACAGACTCATATATATTCCAAAGAGTGAATGCTGCTCCT 1856
DB 1615 AAGTGTACTTCTCCCAACACAGACTCATATATATTCCAAAGAGTGAATGCTGCTCCTC 1674
QY 1857 TATTTTCAGCTTTGCAAGATGAGCCAAAGAGAGTGAAGTGGCTATAGTGACAAATCCA 1916
DB 1675 TATTTTCAGCTTTGCAAGATGAGCCAAAGAGAGTGAAGTGGCTATAGTGACAAATCCA 1734
QY 1917 ATAAATTCAGATGCTTCTATTTGCCAAGAGTGGAAACCATCCACCTCATTAATGCAGT 1976
DB 1735 ATAAATTCAGATGCTTCTATTTGCCAAGAGTGGAAACCATCCACCTCATTAATGCAGT 1794
QY 1977 GGAATGAATGCCACCAATACATTTTAACTCTCAGTTCCTGAAATGACAAAGATGAAG 2036
DB 1795 GGAATGAATGCCACCAATACATTTTAACTCTCAGTTCCTGAAATGACAAAGATGAAG 1854

QY 2037 TTCAAAATGCAGATACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACG 2096
DB 1855 TTCAAAATGCAGATACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACG 1914
QY 2097 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 2156
DB 1915 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 1974
QY 2157 CTTTTTCAGATATTTCCCAAGAAACCGTTCCTCCGAAAGTTTAAAGACATGATGTTAAC 2216
DB 1975 CTTTTTCAGATATTTCCCAAGAAACCGTTCCTCCGAAAGTTTAAAGACATGATGTTAAC 2034
QY 2217 TCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCGTAAATTTTCACTCTTCCA 2276
DB 2035 TCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCGTAAATTTTCACTCTTCCA 2094
QY 2277 AAGAGCCCGAGTGTGATCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2336
DB 2095 AAGAGCCCGAGTGTGATCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2154
QY 2337 TTGAAAGGATACAAATTTGTCCAAGTCAGCCTCTGAGATCAATTTCTGATGAACCTCACAG 2396
DB 2155 TTGAAAGGATACAAATTTGTCCAAGTCAGCCTCTGAGATCAATTTCTGATGAACCTCACAG 2214
QY 2397 CATGCTAAAAATAAAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGGTGGCT 2456
DB 2215 CATGCTAAAAATAAAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGGTGGCT 2274
QY 2457 CCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCCTCCAAACAGCTTAGAGTGTCTGAAAGA 2516
DB 2275 CCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCCTCCAAACAGCTTAGAGTGTCTGAAAGA 2334
QY 2517 TTGCAGAGGTTGACTTTTCTGCAAGTGAAGTGAATAAAGTGAATGCTCATGACCAAGGCTCAG 2576
DB 2335 TTGCAGAGGTTGACTTTTCTGCAAGTGAAGTGAATAAAGTGAATGCTCATGACCAAGGCTCAG 2394
QY 2577 AATCCACCCCTGGAGACCAAGCAAGATTTAGAGTGGAAACTCACACCCCAAAA 2636
DB 2395 AATCCACCCCTGGAGACCAAGCAAGATTTAGAGTGGAAACTCACACCCCAAAA 2454
QY 2637 ACCATAGGCGGAAATGTGACAAAGAAAGCCCATCTCTGATGCTCCACCTGGAAGAGC 2696
DB 2455 ACCATAGGCGGAAATGTGACAAAGAAAGCCCATCTCTGATGCTCCACCTGGAAGAGC 2514
QY 2697 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAGAACAGTAGTAGAGGAA 2756
DB 2515 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAGAACAGTAGTAGAGGAA 2574
QY 2757 AATGCTGAAATCACATAGGCTTACTGAAGTGTACTTGGAGAAAGCTGACAGATTAAC 2816
DB 2575 AATGCTGAAATCACATAGGCTTACTGAAGTGTACTTGGAGAAAGCTGACAGATTAAC 2634
QY 2817 ACAGATAGTTTACTTGGGCTTTTGGCATGGGAAAAAAGTATTTCCAAAGATCTTCTC 2876
DB 2635 ACAGATAGTTTACTTGGGCTTTTGGCATGGGAAAAAAGTATTTTCTAGATCTTCTC 2694
QY 2877 GAGCAAGAGATCATTTGAAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGG 2936
DB 2695 GAGCAAGAGATCATTTGAAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGG 2754
QY 2937 AGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGC 2996
DB 2755 AGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGC 2814
QY 2997 AAGTTTGGATTCATCGGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 3056
DB 2815 AAGTTTGGATTCATCGGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 2874
QY 3057 GTTATCCAGAGCTCAAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTCCAAA 3116
DB 2875 GTTATCCAGAGCTCAAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTCCAAA 2934
QY 3117 GTGGCCATCTGAGGATATGAGCTTTGGCTTCTCTTATTTTATTTATCTCATGATGCA 3176

Db	2935	GTGGCCCAATCTGAGGATATGCAATTTGGCTTCTCTATTTTATATCATGAGTGCA	2994	Db	115	AGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATTGCTGGAAGTCC	174
Qy	3177	GTGCAGGCACATGAATATATCTCAAGTCTTTTGAAGAGTTGATACAGATCAATCTGGTGTC	3236	Qy	357	TTTCAGAAATCGGCTTGTCTGCCAATGCGAATGAGCTGTTGTTTACCTGGGTGAATGGC	416
Db	2995	GTGCAGGCACATGAATATATCTCAAGTCTTTTGAAGAGTTGATACAGATCAATCTGGTGTC	3054	Db	175	TTTCAGAAATCGGCTTGTCTGCCAATGCGAATGAGCTGTTGTTTACCTGGGTGAATGGC	234
Qy	3237	TTGTCTGACAGAGAAATCCGAACTGGCTACAGAAATTCACGAACTCCGCTTAAGTTTG	3296	Qy	417	ACAGATCTTTGACTACTCAAGGAACTACAGCAGGTCAGAGAAACAGATGAGAGAGAGCAG	476
Db	3055	TTGTCTGACAGAGAAATCCGAACTGGCTACAGAAATTCACGAACTCCGCTTAAGTTTG	3114	Db	235	ACAGATCTTTGACTACTCAAGGAACTACAGCAGGTCAGAGAAACAGATGAGAGAGAGCAG	294
Qy	3297	CAGGATTTGACAGGCTCGGAACACATGTAATAAATTCCTCAAAAATGCTTCTCGCTGAT	3356	Qy	477	AAAGCAATGAGAGAAATCTTTGGGAAACACAAACGGAACCTTACTAAGAGAGTGAAG	536
Db	3115	CAGGATTTGACAGGCTCGGAACACATGTAATAAATTCCTCAAAAATGCTTCTCGCTGAT	3174	Db	295	AAAGCAATGAGAGAAATCTTTGGGAAACACAAACGGAACCTTACTAAGAGAGTGAAG	354
Qy	3357	ATCAGCGAGCTAAATATATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3416	Qy	537	CAGTTAGAGTGTCTTAACACACTGCAATTAAGGTGCCAATGCTTGTCTCGACCCAGCC	596
Db	3175	ATCAGCGAGCTAAATATATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3234	Db	355	CAGTTAGAGTGTCTTAACACACTGCAATTAAGGTGCCAATGCTTGTCTCGACCCAGCC	414
Qy	3417	CGGTCACCTAAAGTCTAGTAACAAACCTGTAAACCACTGTAACCACTGACAAAATCCCAAGCA	3476	Qy	597	CTGCCAGCAACATCACCTCTGAAGGACGTGCCATCTCTTTATCTCTTTTCAATCTGCC	656
Db	3235	CGGTCACCTAAAGTCTAGTAACAAACCTGTAAACCACTGTAACCACTGACAAAATCCCAAGCA	3294	Db	415	CTGCCAGCAACATCACCTCTGAAGGACGTGCCATCTCTTTATCTCTTTTCAATCTGCC	474
Qy	3477	TATAAGGACAAAACAAATATAGTTTGAATCATGAGGAGAAAGAAATCGCTTTTAA	3536	Qy	657	AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCTCAGTTGTTGT	716
Db	3295	TATAAGGACAAAACAAATATAGTTTGAATCATGAGGAGAAAGAAATCGCTTTTAA	3354	Db	475	AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCTCAGTTGTTGT	534
Qy	3537	ATGATTCGTACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAACCCCT	3596	Qy	717	TTTGCAGTACTTAAGGATGTTGAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGA	776
Db	3355	ATGATTCGTACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAACCCCT	3414	Db	535	TTTGCAGTACTTAAGGATGTTGAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGA	594
Qy	3597	AGGAAGTTTGTTCCTGGAATGACAAACATTCACCAACATCATAAAGATGCTCAGACAGTG	3656	Qy	777	CAGACAGTATGAGGGGTACTTGAACAACAGATAAGAAAGTCCCTGGATTAGTCTAATG	836
Db	3415	AGGAAGTTTGTTCCTGGAATGACAAACATTCACCAACATCATAAAGATGCTCAGACAGTG	3474	Db	595	CAGACAGTATGAGGGGTACTTGAACAACAGATAAGAAAGTCCCTGGATTAGTCTAATG	654
Qy	3657	AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTG	3716	Qy	837	CAAGATTTTGGCTTCTCCTCAGTGGATTCCACCAACATTCAGGAAACAAATCAACTAAA	896
Db	3475	AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTG	3534	Db	655	CAAGATTTTGGCTTCTCCTCAGTGGATTCCACCAACATTCAGGAAACAAATCAACTAAA	714
Qy	3717	CCAGAGAGTATCAAAACCGTTTCTTCATATGATGATGAGTGCAGGAATGAGGGCTTAT	3776	Qy	897	ACAAATTTGCCAGAAAAATCTTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC	956
Db	3535	CCAGAGAGTATCAAAACCGTTTCTTCATATGATGATGAGTGCAGGAATGAGGGCTTAT	3594	Db	715	ACAAATTTGCCAGAAAAATCTTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC	774
Qy	3777	CGAGACAAATTTGAAGTTTGGAC	3799	Qy	957	AGTGTAGGCTTCTAAAACCTGAAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT	1016
Db	3595	CGAGACAAATTTGAAGTTTGGAC	3617	Db	775	AGTGTAGGCTTCTAAAACCTGAAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT	834
RESULT 12				Qy	1017	AAAGAAACATGACCATTTGATGGAAAAAGAACTGACCAATAGTCTCTGCATATTTATATGG	1076
AR442838				Db	835	AAAGAAACATGACCATTTGATGGAAAAAGAACTGACCAATAGTCTCTGCATATTTATATGG	894
LOCUS	AR442838	3621 bp	DNA	linear	PAT	20-FEB-2004	
DEFINITION	Sequence 21 from patent US 6670165.						
ACCESSION	AR442838						
VERSION	AR442838.1	GI:42670314					
KEYWORDS							
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	1 (bases 1 to 3621)						
AUTHORS	Canfield, W.M.						
TITLE	Methods for producing highly phosphorylated lysosomal hydrolases						
JOURNAL	Patent: US 6670165-A 21 30-DEC-2003;						
FEATURES	Location/Qualifiers						
source	1..3621						
	/organism="unknown"						
	/mol_type="genomic DNA"						
ORIGIN							
Query Match	62.3%; Score 3485.4; DB 6; Length 3621;						
Best Local Similarity	99.7%; Pred. No. 0;						
Matches 3492; Conservative	0; Mismatches 11; Indels 0; Gaps 0;						
Qy	297	AGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATTGCTGGAAGTCC	356				

Db 1195 CTAATGATGATGTCATGTTGGGAAGGATGCTGCGCCAGATGATTTTTTACAGTCACTCC 1254
Qy 1437 AAAGCCAGAGGTTTATTTTACATGSCCTGTGCGAACTGTGCGAGGGTCCCGAGGT 1496
Db 1255 AAAGCCAGAGGTTTATTTGACATGSCCTGTGCGAACTGTGCGAGGGTCCCGAGGT 1314
Qy 1497 TCCTGGATTAGGATGGCTATTGTGACAAGGCTTGTATAATTGAGCCTGCGATTGGGAT 1556
Db 1315 TCCTGGATTAGGATGGCTATTGTGACAAGGCTTGTATAATTGAGCCTGCGATTGGGAT 1374
Qy 1557 GGTGGGATTTGCTCTGGAACACAGTGGAGGGAGTCTGTATATTGAGGAGGTGGAGTACT 1616
Db 1375 GGTGGGATTTGCTCTGGAACACAGTGGAGGGAGTCTGTATATTGAGGAGGTGGAGTACT 1434
Qy 1617 GGGAGTATTGGAGTTGGACACCCCTGGCGAGTTTGGTGGAGGAATAAACAAGTCTCTTTAC 1676
Db 1435 GGGAGTATTGGAGTTGGACACCCCTGGCGAGTTTGGTGGAGGAATAAACAAGTCTCTTTAC 1494
Qy 1677 TGTAAATCAGGATGTGCGAATTTCTTGCTGCTGATAAAGTTTCTGTGACCAAGCATGCAAT 1736
Db 1495 TGTAAATCAGGATGTGCGAATTTCTTGCTGCTGATAAAGTTTCTGTGACCAAGCATGCAAT 1554
Qy 1737 GTCCTTGCTGCTGGGTTTGTATGCTGGCGACTGTGGCGAGATCATTTTTCATGAATTTGAT 1796
Db 1555 GTCCTTGCTGCTGGGTTTGTATGCTGGCGACTGTGGCGAGATCATTTTTCATGAATTTGAT 1614
Qy 1797 AAAAGTGATCCTTCTCCCAAAACAGACTCACTATATTTTCCAAAAGGTGAATGCTGCGCT 1856
Db 1615 AAAAGTGATCCTTCTCCCAAAACAGACTCACTATATTTTCCAAAAGGTGAATGCTGCGCT 1674
Qy 1857 TATTTACGCTTTGAGAGTAGCCAAAGAGGAGTTGAAGTGCTTATAGTGACAATCCA 1916
Db 1675 TATTTACGCTTTGAGAGTAGCCAAAGAGGAGTTGAAGTGCTTATAGTGACAATCCA 1734
Qy 1917 ATAATTCAGACTGCTTCTATTTGCCAAAGTGGAAACCATCCACCTCATATATGCAAGT 1976
Db 1735 ATAATTCAGACTGCTTCTATTTGCCAAAGTGGAAACCATCCACCTCATATATGCAAGT 1794
Qy 1977 GGAATGAATGCCACCAATACATTTTAACTCAGCTTTCAAAATACAAACGATGAAGAG 2036
Db 1795 GGAATGAATGCCACCAATACATTTTAACTCAGCTTTCAAAATACAAACGATGAAGAG 1854
Qy 2037 TTCAAAATTCAGATTAACAGTGGAGTGACACAGGGAGGGACCAAACTGAATCTACG 2096
Db 1855 TTCAAAATTCAGATTAACAGTGGAGTGACACAGGGAGGGACCAAACTGAATCTACG 1914
Qy 2097 GCCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATAAACACTTCTCCAGAGGCGGAAATC 2156
Db 1915 GCCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATAAACACTTCTCCAGAGGCGGAAATC 1974
Qy 2157 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCGAAAGTTTAAGAGACATGATTAAAC 2216
Db 1975 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCGAAAGTTTAAGAGACATGATTAAAC 2034
Qy 2217 TCAACAAGGAGGCGCCAGAGAGGTGAATAATTCCTGTTAAATATTTCATCTCTTCCA 2276
Db 2035 TCAACAAGGAGGCGCCAGAGAGGTGAATAATTCCTGTTAAATATTTCATCTCTTCCA 2094
Qy 2277 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAACATGGAGACATCACT 2336
Db 2095 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAACATGGAGACATCACT 2154
Qy 2337 TTGAAGGATCAATTTGTCGAAGTCAGCCTTGCTGAGATCATTTCTGATGAATCTCAG 2396
Db 2155 TTGAAGGATCAATTTGTCGAAGTCAGCCTTGCTGAGATCATTTCTGATGAATCTCAG 2214
Qy 2397 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGCT 2456
Db 2215 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGCT 2274
Qy 2457 CCACAGGAAAACAGGTTTCAAAAAGATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2516
Db 2275 CCACAGGAAAACAGGTTTCAAAAAGATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2334

Qy 2517 TTGCAGAGGTTGACTTTTCTCAGTGTAGTGTAAAGTGAATGGTTCATGACAGGGTCTAG 2576
Db 2335 TTGCAGAGGTTGACTTTTCTCAGTGTAGTGTAAAGTGAATGGTTCATGACAGGGTCTAG 2394
Qy 2577 AATCCACCCCTGAGCTTTGGAGACCAACAGCAAGATTITAGAGTGGAAACTCACACCCAAAAA 2636
Db 2395 AATCCACCCCTGAGCTTTGGAGACCAACAGCAAGATTITAGAGTGGAAACTCACACCCAAAAA 2454
Qy 2637 ACCATAGGGGAATGTGCAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGC 2696
Db 2455 ACCATAGGGGAATGTGCAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGC 2514
Qy 2697 CAGATCAGAAAAGAAAAGAAAATCACAGGAAAAGAAAAGAGAACAGTAGAATTGAGGAA 2756
Db 2515 CAGATCAGAAAAGAAAAGAAAATCACAGGAAAAGAAAAGAGAACAGTAGAATTGAGGAA 2574
Qy 2757 AATGCTGAAAATCACATAGGCGTTTACTGAAAGTGTACTTGGAGAAAAGCTGACGATTTAC 2816
Db 2575 AATGCTGAAAATCACATAGGCGTTTACTGAAAGTGTACTTGGAGAAAAGCTGACGATTTAC 2634
Qy 2817 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAAGATATTTTCCAAGATCTTCTC 2876
Db 2635 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAAGATATTTTCTAGATCTTCTC 2694
Qy 2877 GACGAAAGAGTCAATTGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATACTGGG 2936
Db 2695 GACGAAAGAGTCAATTGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATACTGGG 2754
Qy 2937 AGGCAACTAAAGATACATTTCCAGATTTCCCTCAGATATGTAATAAATAATTTCTAATAGC 2996
Db 2755 AGGCAACTAAAGATACATTTCCAGATTTCCCTCAGATATGTAATAAATAATTTCTAATAGC 2814
Qy 2997 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCACATGATGACCGGAT 3056
Db 2815 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCACATGATGACCGGAT 2874
Qy 3057 GTTATGCAAGAACTGCAAGATATTTCCCTGGAAGATTTTGAACAGAGCTCATTTACAAA 3116
Db 2875 GTTATGCAAGAACTGCAAGATATTTCCCTGGAAGATTTTGAACAGAGCTCATTTACAAA 2934
Qy 3117 GTCCGCACTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 3176
Db 2935 GTCCGCACTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 2994
Qy 3177 GTCCGCACTGATATATCTCAAGCTTTTGAAGTTGATACAGATCAATCTGGTGTC 3236
Db 2995 GTCCGCACTGATATATCTCAAGCTTTTGAAGTTGATACAGATCAATCTGGTGTC 3054
Qy 3237 TTCTCTGACAGAGAAATCCGACACTGGCTACCGAATTTTCAAGAACTGCGGTTAAGTTG 3296
Db 3055 TTCTCTGACAGAGAAATCCGACACTGGCTACCGAATTTTCAAGAACTGCGGTTAAGTTG 3114
Qy 3297 CAGGATTTGACAGGCTGGAACACATGCTTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3356
Db 3115 CAGGATTTGACAGGCTGGAACACATGCTTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3174
Qy 3357 ATCAGCAGCTAAATAATTTCCACCAACTCAGGATCTCTACTATGATCCCACTGCCA 3416
Db 3175 ATCAGCAGCTAAATAATTTCCACCAACTCAGGATCTCTACTATGATCCCACTGCCA 3234
Qy 3417 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCACAAAGCA 3476
Db 3235 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCACAAAGCA 3294
Qy 3477 TATAAGGACAAAACAAATATAGGTTTGAATCATCGGAGAGAGAAATCGGTTTTTAA 3536
Db 3295 TATAAGGACAAAACAAATATAGGTTTGAATCATCGGAGAGAGAAATCGGTTTTTAA 3354
Qy 3537 ATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAACCCCT 3596
Db 3355 ATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAACCCCT 3414

QY	3362	GCAGCTAAATATATTTCACCAACCTCAGGAATCTCTACTATGATCCCAACCTGCCACCGGT	3421	DB	1981	AGATGTTATTTAAATTAATGACAGCAATATCACCTCTTATTGACAAATACCTAAATTAAGAT	2040
DB	901	GCAGCTAAATATATTTCACCAACCTCAGGAATCTCTACTATGATCCCAACCTGCCACCGGT	960	QY	4502	TTTATTTAAATTTAAAGACTGTAATGGTCTTAAACCACTAACTACTCTGAAGAGCTCAATGA	4561
QY	3422	CACCTAAAGTCTAGTAACAAACTGTAAACAGTAGTAAGTACGACAAAATCCCAAGGATATAA	3481	DB	2041	TTTATTTAAATTTAAAGACTGTAATGGTCTTAAACCACTAACTACTCTGAAGAGCTCAATGA	2100
DB	961	CACCTAAAGTCTAGTAACAAACTGTAAACAGTAGTAAGTACGACAAAATCCCAAGGATATAA	1020	QY	4562	TTGACATCTGAAATGCTTTGTAATTAATGACTTCAGCCCTCAAGAAATGCTATGATTTTCA	4621
QY	3482	GGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATATGAT	3541	DB	2101	TTGACATCTGAAATGCTTTGTAATTAATGACTTCAGCCCTCAAGAAATGCTATGATTTTCA	2160
DB	1021	GGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATATGAT	1080	QY	4622	GTGACAGTCTAATTTCAACAGGCTAGAGTTAGTACTTACCTTACCAGATGTAATTAATGTTTT	4681
QY	3542	TGTCACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGAA	3601	DB	2161	GTGACAGTCTAATTTCAACAGGCTAGAGTTAGTACTTACCTTACCAGATGTAATTAATGTTTT	2220
DB	1081	TCGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGAA	1140	QY	4682	GGAAATGTACATATTCAAAACAGAAAGTCCCTCATTTTGTAGAAAATGAGTAGTCTGATGGCAC	4741
QY	3602	GTTTGTTCCTGTAATGACAACTTGACCAATCATTAAGATGCTCAGACAGTGAAGGC	3661	DB	2221	GGAAATGTACATATTCAAAACAGAAAGTCCCTCATTTTGTAGAAAATGAGTAGTCTGATGGCAC	2280
DB	1141	GTTTGTTCCTGTAATGACAACTTGACCAATCATTAAGATGCTCAGACAGTGAAGGC	1200	QY	4742	TGSCACATTCACAGTGGTGTCTTTTAAATCTCAATCTGGTATATTCAGTAGTATCTCTC	4801
QY	3662	TGTTCTCAGGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAG	3721	DB	2281	TGSCACATTCACAGTGGTGTCTTTTAAATCTCAATCTGGTATATTCAGTAGTATCTCTC	2340
DB	1201	TGTTCTCAGGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAG	1260	QY	4802	TCAGTTGGTTTTTGTATAGAAC--AGAGGCCAGCAAACTTCTTTGTAAAAGGCTGGTAG	4859
QY	3722	AGAGTATCGAAACCGTTTCTTCATATGATGAGCTGCAGGAATGGAGGGCTTATCGAGA	3781	DB	2341	TCAGTTGGTTTTTGTATAGAACAGAGAGGCCAGCAAACTTCTTTGTAAAAGGCTGGTAG	2400
DB	1261	AGAGTATCGAAACCGTTTCTTCATATGATGAGCTGCAGGAATGGAGGGCTTATCGAGA	1320	QY	4860	TAAATTAATTCAGCGCCACCTGTCTTTGTTCATACATCTTCTTGTCTGTTGTTTAGTTTG	4919
QY	3782	CAAAATTGAAGTTTTGCAACCTTGATGTAAGTACACATTTGATTTTACTATATCTC	3841	DB	2401	TAAATTAATTCAGCGCCACCTGTCTTTGTTCATACATCTTCTTGTCTGTTGTTTAGTTTG	2460
DB	1321	CAAAATTGAAGTTTTGCAACCTTGATGTAAGTACACATTTGATTTTACTATATCTC	1380	QY	4920	TTTTTTTTTCAAAACACCTCTTAAAAATGTAAAACCATGTTTAGCTTGACGTGTACAAA	4979
QY	3842	ATTTTTTGTGAGCAGTTAAATTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGATACA	3901	DB	2461	TTTTTTTTTCAAAACACCTCTTAAAAATGTAAAACCATGTTTAGCTTGACGTGTACAAA	2520
DB	1381	ATTTTTTGTGAGCAGTTAAATTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGATACA	1440	QY	4980	AACTGCCACCCAGCCAGATGTGACCTCAGGCCATCATTTGCCAATCACTGAGAAATTA-T	5038
QY	3902	CAAAAGAGTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTGAAAACCATCTAC	3961	DB	2521	AACTGCCACCCAGCCAGATGTGACCTCAGGCCATCATTTGCCAATCACTGAGAAATTA	2580
DB	1441	CAAAAGAGTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTGAAAACCATCTAC	1500	QY	5039	TTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	5098
QY	3962	CTCAGCATTTACTGAGCATTATTTAAAACTCAGCTTCCAGAGATGCTTTGTGATGTGATG	4021	DB	2581	TTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2640
DB	1501	CTCAGCATTTACTGAGCATTATTTAAAACTCAGCTTCCAGAGATGCTTTGTGATGTGATG	1560	QY	5099	AGTGCAAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCCGGGTTCAAGCAGTTCTG	5158
QY	4022	CTTAGAGTTTGGCCGGAAGAGAAATATCCAGTACATGCTGTTTGTGGCATGAT	4081	DB	2641	AGTGCAAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCCGGGTTCAAGCAGTTCTG	2700
DB	1561	CTTAGAGTTTGGCCGGAAGAGAAATATCCAGTACATGCTGTTTGTGGCATGAT	1620	QY	5159	TCTCAGCCTTCTCAGTAGCTGGGACTACAGGTGCAATGCCAACCAACCCCTGCTAAATTTTG	5218
QY	4082	ATAGCCCACTGACTAGGAATTTTAAACCAACCACTGAAATCTGCTGTCGAGCAGCT	4141	DB	2701	TCTCAGCCTTCTCAGTAGCTGGGACTACAGGTGCAATGCCAACCAACCCCTGCTAAATTTTG	2760
DB	1621	ATAGCCCACTGACTAGGAATTTTAAACCAACCACTGAAATCTGCTGTCGAGCAGCT	1680	QY	5219	TATTTTGTAGTAGAGAGGGGGTTCCACCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	5278
QY	4142	CTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTTTTATAAAAG	4201	DB	2761	TATTTTGTAGTAGAGAGGGGGTTCCACCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2820
DB	1681	CTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTTTTATAAAAG	1740	QY	5279	TCAGGTGATCCACTGCTCTGCTCCCAAGTGTGAGATTACAGGCATACAGCCAGTGC	5338
QY	4202	GCTCACTGACAGACAGCTGTTTAAATTTCCACAGCAATCATTTGACACTAACTTTAT	4261	DB	2821	TCAGGTGATCCACTGCTCTGCTCCCAAGTGTGAGATTACAGGCATACAGCCAGTGC	2880
DB	1741	GCTCACTGACAGACAGCTGTTTAAATTTCCACAGCAATCATTTGACACTAACTTTAT	1800	QY	5339	ACCCAGCCGAGAAATTTAGTATTTTATGTTTAAACCTTTGGCGTCTAGCCATATTTTA	5398
QY	4262	AGGAGAAGCCTATGCCAGCTGGGAGTGTGTAAGAGGCTCCAGTCTTTGCATTTCCAAA	4321	DB	2881	ACCCAGCCGAGAAATTTAGTATTTTATGTTTAAACCTTTGGCGTCTAGCCATATTTTA	2940
DB	1801	AGGAGAAGCCTATGCCAGCTGGGAGTGTGTAAGAGGCTCCAGTCTTTGCATTTCCAAA	1860	QY	5399	TGTCATAATACAATGAGTTTGTGAAGAGCAGATTCCATGAGTAACTCTGACAGTATTTT	5458
QY	4322	GCCTTTTGTAAAGTTTGGCATTTTATTTTCCATTTTCCCATTTTAAAGTAGTACTAA	4381	DB	2941	TGTCATAATACAATGAGTTTGTGAAGAGCAGATTCCATGAGTAACTCTGACAGTATTTT	3000
DB	1861	GCCTTTTGTAAAGTTTGGCATTTTATTTTCCATTTTCCCATTTTAAAGTAGTACTAA	1920	QY	5459	AGATCATGATCTCAACAAATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGG	5518
QY	4382	GTTAACTAGTTATCTTGCTTCTGAGTATTAAGAAATTTGGGATGCTTAAACCTATTTTAT	4441	DB	3001	AGATCATGATCTCAACAAATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGG	3060
DB	1921	GTTAACTAGTTATCTTGCTTCTGAGTATTAAGAAATTTGGGATGCTTAAACCTATTTTAT	1980	QY	5519	AAATGTAATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	5578
QY	4442	AGATGTTATTTAAATATGACAGCAATATCACCTCTTATTGACAAATCACTAAATTAAGAT	4501				

Db 3061 AAGTAACTACTGTGTTGGTGTAGAGTTGTGTATTTCAAAAACTGAATCTCAT 3120

QY 5579 AAAAGTAAATTTGAAAA 5597

Db 3121 AAAAGTAAATTTGAAAA 3139

RESULT 14

AKI73132

LOCUS AKI73132 5204 bp mRNA linear ROD 28-JUL-2004

DEFINITION Mus musculus mRNA for mKIAA1208 protein.

ACCESSION AKI73132

VERSION AKI73132.1 GI:50510848

KEYWORDS FLI_CDNA.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE

AUTHORS Okazaki, N., Kikuno, R.F., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S., Saga, Y., Seino, S., Nishimura, M., Kaisho, T., Hoshino, K., Kitamura, H., Nagase, T., Ohara, O. and Koga, H.

TITLE Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene: IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries

JOURNAL DNA Res. 11, 205-218 (2004)

REFERENCE

AUTHORS Okazaki, N., Kikuno, R.F., Nagase, T., Ohara, O. and Koga, H.

TITLE Direct Submission

JOURNAL Submitted (19-MAY-2004) Hisaaki Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kenatari, Kisarazu, Chiba 292-0818, Japan (E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)

COMMENT The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.

FEATURES

Location/Qualifiers

1..5204

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="mbg11133"

/tissue_type="brain"

/dev_stage="adult"

/note="vector:modified pBC SK+"

1..5204

/gene="mKIAA1208"

<1..3715

/gene="mKIAA1208"

/note="CDS is predicted by in silico analysis. Start codon is not identified."

/codon_start=2

/evidence=not_experimental

/product="mKIAA1208 protein"

/protein_id="BAD32410.1"

/db_xref="GI:50510849"

/translation="GVMLLKLRQTYCLSHRYGLVCFVGVVVTVISAFQFGEVVL EWSRDYHVLGRVDNIAKGSFONRLCLPMDIVVTVWNGFDLLELQOVRHMF EEOARARETLGKNTTPTKSKSEKLECLTHCIKYPMLVLDPLPANCITLKLPLY EAPGLVMQGLAFUSGPPFTFKTSQKLKLPKLSKIKLLRXYSEASVALLKLNRP XGFELNKTQKNTIDGKELTISPAVLLMDLSAISQSKODEVSASFEDNEELYS LRSIERHAPWRNFIIVTNGQIIPSWLNDNPRVITVHQDIFQNLHLPTFSSPALES HIHRIEGLSQFVILNDVMFGDVPDPDFYSHSKGVKQVYLTPWPNCAEGCPGSIK DGYCDKACNNSADGDCGSGTAGNRFVAGGGTGNIGAGOHQFQGGGINTISVCN OGCSANSLADKFCDDQACNVLSCGFDAGDCGDHFELYKVTLIPNTHYVVPKEVLS YFSFANTARGVEGTSNDPIIRHASIANRWKTIHILHMSGMNATIYFNLIQNAND EEFKIQIAVEVDTRKAPLNSTTKQAYESLVSPVLPQADVPFDVPPKPKPKLR HDVNATRFQEEVKIPRNISLIPKAEQVLSNLDLQLERGDIITLKGYNLSALLRS FLGNSLDTKI QPQARTKGNLEVPQENSHRPHGPAGHRSERWPAETVTVKG RDHALNPPVPLETNARLAQPTLGVTVSKENLSPLIVPPSSHLPKKEESDRAEGNAVPT

ORIGIN

Query Match 51.4%; Score 2874.6; DB 10; Length 5204;

Best Local Similarity 79.6%; Pred. No. 0;

Matches 3681; Conservative 0; Mismatches 814; Indels 128; Gaps 19;

QY 159 GGGGTGATGCTGTTCAAGCTCTCGAGAGACAACTATACCTGCTGCTGCCAGGTAT 218

Db 2 GGGGTGATGCTGTTCAAGCTCTCGAGAGACAACTATACCTGCTGCTGCCAGGTAT 61

QY 219 GGGCTCTAGTGTGCTTCTTGGGGCTGCTGTGCACCTGCTCGCGCTTCCAGTTCGA 278

Db 62 GGGCTCTAGTGTGCTTCTTGGGGCTGCTGTGCACCTGCTCGCGCTTCCAGTTCGA 121

QY 279 GAGTGGTCTGGAATGAGCGGAGATCAATACCAATGTTTGTGATCTCTATAGAC 338

Db 122 GAGTGGTCTGGAATGAGCGGAGATCAGTACCATGTTTGTGATCTCTATAGAC 181

QY 339 AATATTGCTGGAAGCTCTTTCAGAACTCGGCTTGTCTGCCATGCCGATGAGTTCGT 398

Db 182 AATATTGCTGGAAGCTCTTTCAGAACTCGGCTTGTCTGCCATGCCGATGAGTTCGT 241

QY 399 TACACTGGTGAATGGCACAGATCTTGAACCTACTGAAAGAACTACACAGGTGAGAA 458

Db 242 TACACTGGTGAATGGCACCTGACCTTGAACCTGCTTAAAGAGCTACACAGGTGAGAA 301

QY 459 CAGATGGAGGAGGACAGAAAGCATGAGAAATCTTGGGAAACACACACAGGACCT 518

Db 302 CACATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361

QY 519 ACTAAGAGAGGTGAGAGAGCAGTATAGAGTGTGTTGCTTAAACACTGATTAAGGTGCCAATG 578

Db 362 ACTAAGAGAGGTGAGAGAGCAGTATAGAGTGTGTTGCTTAAACACTGATTAAGGTGCCAATG 421

QY 579 CTTGTACTGACCCAGCCCTGCCAGCAACATCACCCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 638

Db 422 CTTGTACTGACCCAGCCCTGCCAGCAACATCACCCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 481

QY 639 CTTCTCTTTCATCTGCGCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTCTTACC 698

Db 482 CCACTCTTCCAGCTGCCAGCAACATGTTCAATGTTGCAAAACCAAAACCCCTCTTACC 541

QY 699 AATGTCTCAGTGTGTTTGTGAGTACTAGAGATGTTGAAGATGTTGAGATGTTGAGATGTTGAGATG 758

Db 542 AATGTCTCAGTGTGTTTGTGAGTACTAGAGATGTTGAAGATGTTGAGATGTTGAGATGTTGAGATG 601

QY 759 CTTAAGGAATAGCAGACAGCAGTATGAGGGGGTACTTGACACAGATTAAGAGTCT 818

Db 602 TTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 661

QY 819 CTTGATTTAGTCTAATGCAAGATTTGCTTCTGAGTGGATTTTCCACCAACATTTCAAG 878

Db 662 CTTGCTTAGTCTGATGCAAGGCTTGGCGTTTCTGAGTGGATTTTCCACCAACATTTCAAG 721

QY 879 GAAACAATCACTAATAAACAATAATGTCAGAAATCTTCTCTTAAAGTCAAACTGTTG 938

Db 722 GAGACAGTCACTGAAGACAAAGCTGCCAGAAAAGCTTCTCTTAAAGTCAAACTGTTG 781

QY 939 CAGTTGTATTTCAGAGGCGCAGTGTAGCGCTTCTTAAACATGAAATACCAAGGATTTTCAA 998

Db 782 CGGCTGTACTCGAGGCGCAGTGTGCTCTTCTGAAATGTAATATCCCAAGGGTTTCAA 841

QY 999 GAATTGAATAGCAAACTAAGAGAAACATGACCATTTGATGGAAGAACTGACCATAGT 1058

Db 842 GAGCTGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 901

QY 1059 CCTGCATATTTATTTAGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGATC 1118
DB 902 CTTGCGTATCTGCTGTGGAGCTTGAGTCCCATCAGCCAGTCCAAGCAGGATGAGAGCTG 961
QY 1119 TCTGCCAGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCAITTCGGATCTATCGAGAGG 1178
DB 962 TCTGCCAGCGCTTTGAGGATAATGAAGAGCTGAGGTACTCGCTCGATCTATCGAGAGA 1021
QY 1179 CATGCACATGGTTTCGGAATATTTTCAITTCAGCAACGGGCGAGATTCATCTCTGGCTG 1238
DB 1022 CACGGCCATGGGTACGGAATATTTTCAITTCAGCAACGGGCGAGATTCATCTCTGGCTG 1081
QY 1239 AACCTTGACAACTCTCGAGTGACAAATAGTAACACACAGGAGTGTTCGAAATTTGAGC 1298
DB 1082 AACCTTGACAACTCTCGAGTGACAAATAGTAACACACAGGAGTGTTCGAAATTTGAGC 1141
QY 1299 CACTTGCCTACCTTTAGTTACCTCTCTATTTGAAGTCACTTCATCGCATCGAAGGGCTG 1358
DB 1142 CACTTGCCTACTTTTCAGTTCCCTCTCTATTTGAAGTCACTTCATCGCATCGAAGGGCTG 1201
QY 1359 TCCCAAGAGTTTATTTTACTTAATCATGATGTCATGTTTGGAGGAGTGTCTGCGCCAGAT 1418
DB 1202 TCCCAAGAGTTTATTTTACTTAATCATGATGTCATGTTTGGAGGAGTGTCTGCGCCAGAT 1261
QY 1419 GATTTTTCAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAACTGT 1478
DB 1262 GATTTTTCAGCCACTCCAAAGGTCAAAAGGTTTATTTGACATGGCTGTGCCAACTGT 1321
QY 1479 GCCGAGGGCTGCCAGGTTCTCGATTAAGGATGCTATTTGTGACAGGCTTGTAAAT 1538
DB 1322 GCAGAGGGCTGCCGGGCTCTCGATTAAGAGCGGCTATTTGTGTAAGGCTGTAAAT 1381
QY 1539 TCAGCTCGGATTTGGGATTTGGGATTTGCTCTGGAACAGTGGAGGGAGTCCGCTATAT 1598
DB 1382 TCAGCTGTGACTGGGATGGCGGAGCTGCTCTGTTATATCTGAGGGAACGGGTTGTT 1441
QY 1599 GCAGAGGTGGAGTACTGGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGA 1658
DB 1442 GCAGAGGTGGGGAATTTGGAGTTGGACAGCACTGGCAGTTTGGTGGAGGA 1501
QY 1659 ATAAACAGTGTCTTACTGTAAATCAGGAGTGGGAATTCCTGGCTCGCTGATTAAGTTC 1718
DB 1502 ATAAACAGTGTCTTACTGTAAATCAGGAGTGGGAATTCCTGGCTCGCTGATTAAGTTC 1561
QY 1719 TGTGACCAAGCATCAATGTCTTGTCTGTGGGTTTGTGCTGGCAGTGTGGGCAAGAT 1778
DB 1562 TGTGACCAAGCTGTAACTGTCTTATCTCTGGGTTTGTGCTGGTGTGCTGTGGCAAGAT 1621
QY 1779 GATTTTTCATGATTTGATTAAGTGTCTCTCCCAACCCAGACTCACTATATTTATTTCCA 1838
DB 1622 CATTTTCATGAATGTATAAGTAAACACTTCTCCCAACCCAGACTCACTATGTGTGCCCC 1681
QY 1839 AAAGGTGAATGCTGCTTATTTTCAAGTTTGCAGAGTAGCCAAAGAGGAGTTGAAGGT 1898
DB 1682 AAAGGTGAATACCTGTCTTATTTTCAAGTTTGCAGAGTAGCCAAAGAGGAGTTGAAGGT 1741
QY 1899 GCCTATAGTGACAAATCCCAATATTCGACATGCTTCTTATGCAACAGTGGAAACCAATC 1958
DB 1742 ACCTACAGCGACAACCCCATCATCCGCCACGCTCCATTTGCAAAACAGTGGAAACCCATA 1801
QY 1959 CACCTCATATTCACAGTGGATGAATGCCACCAATACATTTTAACTCAGCTTTTCAA 2018
DB 1802 CACCTGATATTCACAGTGGATGAATGCCACCAATACATTTTAACTCAGCTTTTCAA 1861
QY 2019 AATACAAACGATGAAGAGTTCAAAATGCAGATTAACAGTGGAGGTGGACACAGGAGGGA 2078
DB 1862 AACGCCAACGACGAAGAGTTCAAGTTCAGATAGCAGTAGAGTGGACACAGGAGGCGG 1921
QY 2079 CCAAACTGAATCTACGCCCCAGAGGTTACGAAATTTAGTTAGTCCCATTAACCTT 2138
DB 1922 CCAAACTGAATCTTACAAACCCAGAGGCTATGAAGTTTGGTTAGCCCAAGTACACT 1981
QY 2139 CTTCCAGAGCGGGAATCTTTTTCAGGATATTTCCAAAGAAACCGCTTCCCGAAGTTT 2198

DB 1982 CTTCTCAGGCTGACGCTCCTTTTGAAGATGTCCTCCCAAGAGAAACGCTTCCCAAGATC 2041
QY 2199 AAGAGACATGATTTAACTCAACAGAGAGGCCAGGAGAGGTGAAATTTCCCTGGTA 2258
DB 2042 AGGAGACATGATTAATGCAACAGGAGATTTCCAAGAGAGGTGAAATTTCCCGGGTA 2101
QY 2259 AATATTTTCACTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTTGATTTGCAACTG 2318
DB 2102 AATATTTTCACTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTTGATTTGCAACTG 2161
QY 2319 GAACATGGAGACATCACTTTTGAAGGATCAAAATTTGTCCAAGTCAAGCTTGTGTGAGATCA 2378
DB 2162 GAACGTTGGAGACATCACTCTGAAAGGATATACTTGTCCAAGTCAAGCTTGTGTGAGATCA 2221
QY 2379 TTTTCTGATGAACATCAGCATGCTAAATTAATAATCAAGCTTATATAACAGATGAACA 2438
DB 2222 TTTCTGGGGAATTCATAGATATAATAATAAACCCTCAAGCTAGG---ACCAGTGAACA 2278
QY 2439 AATGACAGTTTGGTGGCTCCACAGGAAACACAGTTTATAAAGCATCTTGCACAAACAGC 2498
DB 2279 AAAGGCAACCTGAGGTCCACAGGAAACCCCTTCTCACAGCG---TCCACATGGC 2332
QY 2499 TTAGGAGTGTCTGAAAGATTGACAGGTTGACTTTTCTTGCAGTGTGTAAAGTGAAT 2558
DB 2333 TTTTCTGGTGAACACACAGATCAGAGATGGAATGCTGCCACAGCAGAGCAGTGAACGCTGAAA 2392
QY 2559 GGTTCATGACAGGCTCAGATCCACCCCTGGACTTCGAGACACACAGCAAGATTTAGAGTG 2618
DB 2393 GGCCGTGACCACTTTGATTCACCCCGGTGTGGAGCCATGTCAGATTT----- 2445
QY 2619 GAAACTCACACCCCAACCAATAGCGGAAATGTGACAAAGAAAGCCCATCTCTG 2678
DB 2446 -----GGCCCAAGCTTACCTAGGCTGCTGTGTCCAAAGAGAACCTTTTCCCGCTG 2497
QY 2679 ATTGTTCCACTGAAAGCCAGATGACAAAGAAAGAAATACACAGGGAAGAAAGAG 2738
DB 2498 ATCGTTCCCCCAAGAACCACTTTGCCAAAGAAAGAGGAGAGTGACAGGGCAGAA----- 2551
QY 2739 AACAGTAGAATGGAGAAATGCTGAAATCACATAGGCGTTTACTCAAGTGTACTTTGGA 2798
DB 2552 -----GGCAATGCTGTACTTGTAAAGAGTTAGTGCCTGGC 2587
QY 2799 AGAAGCTGAGCATTTACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAG 2858
DB 2588 AGACGTTGCGAGC-----AGAAATATCCAGGCTTTTGGCCTGGGAGAAAG 2638
QY 2859 TATTTTCAAGATCTTCTCGACGAAGAGAGTCAATGAAGACACAAATTTGGCATACTTCACT 2918
DB 2639 TATTTTCAAGACCTTCTTGTAGGAGAGAGTCAATGAAGACCCAGTTGGGTACTTTACA 2698
QY 2919 GATGACAAATATCTGGAGGCAACTAAAAGATACATTTTGCAGATTCCTTCAGATATGTA 2978
DB 2699 GACAGCAACATACCGGAGGCAACTAAAAGATACATTTGCAGACTCCCTCCGATACGTC 2758
QY 2979 AATTAATTTCTAANTAGCAAGTTTGGATTCACATCCGGAAGTCCCTGCTCACATGCCCT 3038
DB 2759 AATTAATTTCTCAACAGCAAGTTTGGATTCACATCCAGGAAGTCCCTGACACATGCCG 2819
QY 3039 CACATGATTCACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGAC 3098
DB 2819 CACATGATTCACAGGATCGTTATGCAAGAACTCCAAGATATGTTCCCTGGAAGATTTGAC 2878
QY 3099 AAGAGCTATTTCAAAAGTGGCCATTTCTGAGGATATGCGAGTTTGCCTTCTTATTTT 3158
DB 2879 AAGACTTCATTTCAAGGTGGCTCTCTGAGGACATGCAAGTTTGCCTTCTCTACTTT 2938
QY 3159 TATTTATCTATGAGTGCAGTGCAGCCACATGAAATATATCTCAAGTCTTTGATGAAGTTGAT 3218
DB 2939 TATTTATCTATGAGTGCAGTGCAGTGCAGCCCTCAATATTTCCCAAGTCTTTTCAAGTAGAC 2998
QY 3219 ACAGATCAATCTCGTGTCTGTCAGAGAAATCCGAACACTCGGTACCGAATTCAC 3278

Db 2999 ACAGACCAATCTGGTGTCTTCTGTGATAGGGAATCCGAACACTGGCCACGAGAAATTCAC 3058
QY 3279 GAATCGCGTTAAGTTTCAGAGATTGACAGGTCTGGAACACATGCTTAATAAATGCTCA 3338
Db 3059 GACCTACCTTTAAGCTTCAGAGATTGACAGGTTCGGAACACATGTTAATAAATGCTCA 3118
QY 3339 AAAATGCTTCCTGCTGATACAGCAGCTAAATATATATCCACCACTCAGGATCCCTAC 3398
Db 3119 AAAATGCTTCCTGCTGATATCACTCACTCAACCAATCCACCGATCCAGAAAGCATAC 3178
QY 3399 TATGATCCCAACCTGCCACCGGTCTCAATAAAGTCTAGTAAACAACTGTAAACCAAGTAACT 3458
Db 3179 TAGGACCCCAACCTGCCCTCCGCTCACTAAAGTCTTGTGACCAACTGTAAAGCCAGTAACT 3238
QY 3459 GACAAATCCCAAGCATATAAGGACAAACAAATATAGGTTTGAATCATGGGAGAA 3518
Db 3239 GACAAAGATCCCAAGAGCCTATAAAGACAAAGCAAAATACAGGTTTGAATCATGGGAGAG 3298
QY 3519 GAAGAAATCGCTTTTAAATGATTGCTACCAACCGTTTCTCATGTGGTTGGCCAGTTGGAT 3578
Db 3299 GAAGAAATCGCTTTCAAGATGATACGAACCAATGTTTCTCATGTGGTTGGCCAGTTGGAT 3358
QY 3579 GACATAAGAAAACCCCTAGGAAGTTTGTTCCTGCTGAATGACAAATGACCAACATCAT 3638
Db 3359 GACATCAGAAAACCCCGAGGAGTTCTGTTGTCTGAATGACAAATGACCAACCAATCAT 3418
QY 3639 AAAGATGCTCAGACATCAAGCGTCTCTCAGGAGCTTCTATGAATCCATGTTTCCCATTA 3698
Db 3419 AAAGATGCCGACAGTGAAGGCTGTCTCAGGAGCTTCTATGAGTCCATGTTTCCCATTA 3478
QY 3699 CTTTCCCAATTGAATCCCAAGAGGATCGAAACCGTTTCTTCATATGATGATGAGCTG 3758
Db 3479 CTTTCCCAAGTTGAGCTGCCAAGAGGATCGGAACCGCTTCTGCAATGATGAGCTC 3538
QY 3759 CAGGAATGGGGCTTATCGAGACAAATGGAATTTGGACCCATCTGTACTAGCAACA 3818
Db 3539 CAAGAAATGGGGGCTATCGAGACAAAGCTGAAGTTTGGACCCACTCGTACTAGCAACG 3598
QY 3819 TTGATTATGTTTACTATATCTCAATTTTCTGCTGAGCAGTTAATTTGCACTTAAGCGAAG 3878
Db 3599 TTGATTATATTTACTATATCTCAATTTTCTGCTGCAACAGATAATTTCTGAGCGGAAG 3658
QY 3879 ATATTTCCCAAGAGGATACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAG 3938
Db 3659 ATATTTCCCAAGAGGAGGATACAAAGAGCTAGTCCCAAGCAATCAGGGTGTAGAAG 3718
QY 3939 ATCTTCATTTGAAAACCACTTACTCAGCATTTTACTGAGCATTTTAAAGCTCAGCTTCAC 3998
Db 3719 ATCTTCATTTGAAAGTCACTTACTCAGCATCT-GTGAACATCTCCCTCTCGACACCAC 3777
QY 3999 AGAGATGCTTTGTGATGTAGTGTGA-GCAGTTTGGCCCGCAAGAGAAATATCCAGT 4057
Db 3778 AGCGAGTCCCTGTGATGTGGCAGAGGCGAGCCCTCGTGGGAGAGGAGCATCGTGCAG 3837
QY 4058 ACCATGCTGTTTGTGGCATCAATATAGCCACCTGACTAGGAATTTATTAACCAACCCAC 4117
Db 3838 ACCGGGTTCTTCTGCAATGGGAAGAGAGCCCACTGACCTGGAATATT-CAAGACAC 3893
QY 4118 TGAATACTGTGTGTCGAGCAGCTCTGAATCTGATTTTACTTTTAAAGAAATTTGCTCATGG 4177
Db 3894 TAAGAACCTGTGTCAAATAGC---TTGTACAGCTTGTACTTTTAAAGAAATTTGCGGAAGG 3949
QY 4178 ACCTGTCACTTTTATTAAGAGGCTCACTGCAAGAGAGCAGTGTATTTTCCACAG 4237
Db 3950 ACCTGTGCGCTGTGTGAACAAACCTCCCTGACA-----AGTGTGTTTCTTCCCC 4001
QY 4238 CAATCATTTGCAGACTAATTTTATAGGAGAGCCCTATGCCAGCTGGGAGTGAATGCTAAG 4297
Db 4002 CAGTTACTGCAGACT-----GAGAAACAGTCCATCTTTTGAAGCAAGTGGGAG 4050
QY 4298 AGGCTCAGCTCTTTCATTCAGAGCCTTTTGTCTAAAGTTTGTGACATTTT-----TTT 4350
Db 4051 GGGCCCCAGCTCTTTGTCATTCAGAGCTTTTCAGCATATTTCTGGCTGTCTCTCTCTCTCT 4110

QY 4351 TTTTCATTTCCCATTTT-----AAGTAGTTACTAAGTTAACTAGTTATTTCTTGC 4400
Db 4111 GATCCATTTCCCATTTT-----TTTTAAAAAACAATAAGTGGCTACTAAGTTAGTCAITCTCAC 4170
QY 4401 TTCTGAGTATAACGAATTGGGATGTTAAACCTATTTTATAGATGTTTATTTAAATAATG 4460
Db 4171 TTCTCAAAATAACAATCAGGATGTCAAAAC--ATTGTATAGATCTTTATTTAAATATA 4228
QY 4461 CAGCAATATCAGCTCTTATTTAGCAATACCTAAATATGAGTTTATTTAAATTTAAGACT 4520
Db 4229 TAGAAGCATTTACTTCTTT---AGCCTATCTAAATTTATTTTATTTAAACAGTCAAGTG 4285
QY 4521 GTAAATGGTCTTTAAACCACTAACTACTGAAGAGCTCAATGATTGACATCTGAAATGCTTT 4580
Db 4286 GTCTTGAATCGCTAAC-----AACTACTGAAGAGCTCGA-GATTGACGTTGAAAGTCTTT 4340
QY 4581 GTAATTTATGACTTCAGCCCTTAAGAATGCTATGATTTCAAGTGCAGGTCTTAATTTCAAC 4640
Db 4341 GAGCTTGTTTAACTCATTTCCCAAGATACTGTGACCTCGTGTGGCGGCTGATTGCCAA 4400
QY 4641 AGCTAGAGT-----TAGTACTACTTACCAGATGTAATTTATTTTGGAAATGTACAT 4693
Db 4401 GGGCTAGTGTACAGTAGCAGTGTCTCACCAGATGTAATTTATGCTGTGGAAATGTACAT 4460
QY 4694 ATTCAAAACAGAAAGTGCCTCAITTTTAAAGATGAGTAGTGTCTGATGGCACTGGCACATTACA 4753
Db 4461 A--CAACAAAGTGCCTCACTTCAGAAATGAGTAGTGTCTGATGSCACCAGAGTGTATG 4518
QY 4754 GTG 4756
Db 4519 GTG 4521

RESULT 15
AR300468
LOCUS AR300468 5229 bp DNA linear PAT 12-JUN-2003
DEFINITION Sequence 16 from patent US 6537785.
ACCESSION AR300468
VERSION AR300468.1 GI:31687910
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5229)
AUTHORS Canfield, W.M.
TITLE Methods of treating lysosomal storage diseases
JOURNAL Patent: US 6537785-A 16 25-MAR-2003;
FEATURES
source
1..5229
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 50.4%; Score 2820.8; DB 6; Length 5229;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;

QY 150 GSCGCTGAAGGGGTGATGCTGTTTCAAGCTCTCTGCAAGACAAACCTATACCTGCTGCTGCC 209
Db 1 GSCGCTGAAGGGGTGATGCTGTTTCAAGCTCTCTGCAAGACAAACCTATACCTGCTGCTGCC 60
QY 210 CACAGGTATGGGCTCTAGCTGTGCTTCTTGGSCGCTGTTGTCACCATGCTTCGSCCTTC 269
Db 61 CACAGGTATGGGCTCTAGCTGTGCTTCTGTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTC 120
QY 270 CAGTTCCGAGAGGTGGTCTTCTGGAATGGAGCCGAGATCAATACCATGTTTCTGTTGATTC 329
Db 121 CAGTTCCGAGAGGTGGTCTTCTGGAATGGAGCCGAGATCAGTACCATGTTTCTGTTGATTC 180
QY 330 TATAGACAAATATGCTGGAAAGTCTTTCAGAAATCGGCTTTGCTGCCCATGCCGAT 389
Db 181 TACAGAGACAAATGCTGGGAAATCCTTTTCAAGATCGGCTCTGTCTGCCCATGCCAATC 240

Qy	390	GACGTTGTTTACACCTGGGTGAATGGGCACAGATCTTGAACTACTGAAGGAACCTACAGCAG	449
Db	241	GACGTGTTTATACACCTGGGTGAATGGCACTGACCTTGAACTGCTAAAGGAGCTACAGCAG	300
Qy	450	GTCCAGAAACAGATGGAGGAGGAGCAAGACGAATGAGAGAAATCCCTTGGGAAAAACACA	509
Db	301	GTCCGAGAGCACATGGAGGAGAGAGCAGAGACCAATGCCGGAACCCCTCGGGAGAACACA	360
Qy	510	ACGGAACTTACTAAGAGAGGTGAGAACGAGTGTAGAGTTGTGTAAACACTCGATTAAG	569
Db	361	ACCGAACCGAAGAAGAGGTGAGAGCAGCTGGAATGTCTGTGACGCACTGCATTAAAG	420
Qy	570	GTGCCAATGCTGTACTGGACCCAGCCCTGCCAGCCAACTCAACCTGGAAGGACGTGCCA	629
Db	421	GTGCCCATGCTTGTCTGGACCCCGCCCTGCCAGCCACCACTCAACCTGGAAGGATCTGCCA	480
Qy	630	TCTCTTTATCCTCTCTTTTCAATTCGCGAGTGAACATTTTCAATGTTGCCAAACCAAAAAC	689
Db	481	ACCTTTTACCACATCTTTCCACGCGTCCAGCGGACATGTTCAATGTTGCCAAACCAAAAAT	540
Qy	690	CCCTCTACCAATGCTCAGTGTGTGTTTTTGGACAGTACTAAGGATGTTGAAGATGCCAC	749
Db	541	CCGTCTACAATGTCCCGTGTGCTTTTTTGGACACTAAGGATGTTGAAGAGGCCCAT	600
Qy	750	TCTGGACTGCTTAAAGGAATAGCAGACAGACAGTATGGAGGGGTACTTTGCAACAGAT	809
Db	601	GCTGGACCGTTTAAGGAGGCCAGCAACAGATGTTTGGAGAGCCTACTTGAACAACAGAC	660
Qy	810	AAAGAACTCCCTGGATTTAGTGCTAATGCAAGATTTGGCTTTCTCTGAGTGATTTCCACCA	869
Db	661	AAAGACCCCTTGCGCTTAGTGCTGATACAAAGCTTTGGCGTTTCTGTAGTGATTTCCACCG	720
Qy	870	ACATTTCAAGGAAACAAATCAACTAAAAACAAAAATGCCAGAAAACTTTCTCTCT---AAA	926
Db	721	ACCTTTCAAGGAGACGAGTCAACTGAAGACAAAAGCTGCCAAGAAAAGCTTTCCCTCTAAAA	780
Qy	927	GTCAAACGTGTTGAGTTGTATTTCAGAGGCCAGTGTAGCGCTTTCTAAACATGAATACCCC	986
Db	781	ATAAAGCTGTGTGGCGCTGTACTCGAGGCCAGTGTGCGCTCTTTCTGAATTTGAATAATCCC	840
Qy	987	AAGGATTTTCAAGAAATTTGAATTAAGCAAACTAAGAAAGAAATGACCACTGATGAAAAGAA	1046
Db	841	AAGGGTTTCCAAGAGCTGAACAAGCAGCACCAAGAAAGCAATGACCATCGATGGGAGGAA	900
Qy	1047	CTGACCAATAGTCTCTGCATATTTATTTATGGATCTGAGCGCCATCAGCCAGTCTAAGCAG	1106
Db	901	CTGACCAATCAGCCCTGCGTATCTGCTGTGGACCTGAGTCCCATCAGCCAGTCCAAAGCAG	960
Qy	1107	GATGAAGACATCTCTGCCAGTCGTTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGCCGA	1166
Db	961	GATGAGACCGCTCTGCCAGCGCTTTCAGCGCGCTTTGAGGATPAATGAAGAGCTGAGTACTCGCTCGGA	1020
Qy	1167	TCTATCAGAGGCAATGCACATGCGGTTTCGGAATATTTTTCATTTGTCCCAACGGCGCAGATT	1226
Db	1021	TCTATCAGAGACACGGGCCATGGGTACGGAATATTTTCAATTTGTTCACCAACGGCGCAGATT	1080
Qy	1227	CCATCTTGGCTGAACCTTGACAACTCTCGAGTGACAAATAGTAACACACACAGGATGTTTTT	1286
Db	1081	CCATCTTGGCTGAACCTTTGACACCCCTCGAGTGACCAATAGTGACCCACAGGACATTTTC	1140
Qy	1287	CGAAATTTGAGCCACTTTCGCTPACCTTTAGTTCACCTGCTATTTGAAAGTCAATTCATPCGC	1346
Db	1141	CAAAATCTGAGCCACTTTCGCTACTTTTCAGTTTCCCTGCTATTTGAAAGTCAATTCACCCG	1200
Qy	1347	ATCGAAGGGCTGCCCAAGATTTATTTTACTTAATGATGATGTCATGTTTGGGAAGGAT	1406
Db	1201	ATCGAAGGGCTGTCCCAAGATTTATTTATCTAAATGACGATGTCATGTTTCGTTAGGAC	1260
Qy	1407	GTCTGGCCAGATATTTTTCAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCT	1466
Db	1261	GTCTGGCCGACGATTTTTCAGGCCATCCAAAGGTCAAAAGGTTTATTTGACATGGCCT	1320

Qy	1467	GTGCGAAACTGTGTCGCGAGGGCTGCCACAGGTTCTCTGATTAAGGATGGCTATTGTGACAAG	1526
Db	1321	GTGCCAAACTGTGCGAGAGGGCTGCCCGGGCTCTCTGATTAAGGACGGCTATTGTGATAAG	1380
Qy	1527	GCATTGTAATAATTACAGCTCTCGGATTTGGGATGTGGGGAATGCTCTGCGAAACAGTGGAGGG	1586
Db	1381	GCCTGTAAATACCTCAACCTGTGACTGGATGGCGGAAACTGCTCTGGTAATACTGCAGGG	1440
Qy	1587	AGTCGCTATATTCGACGAGAGGTGGAGGTACTTGGGAGTATTGGAGTTTGGACACCCCTGGCAG	1546
Db	1441	AACCGGTTTTTGTGCAAGAGGTGGGGGTACCGGGAATATTGGAGCTGGACAGCACTGGCAG	1500
Qy	1647	TTTGTGTGGAGGAATAAACAGTGTCTCTTACTGTAAATCAGGGAATGTCGAATTTCTGGCTC	1706
Db	1501	TTTGTGTGGAGGAATAAACCACTCTCTTACTGTAAACAAGGATGTGCAAACTCTGGCTG	1560
Qy	1707	GCTGATAAGTTTCTGTGACCAAGCATCAATGTCTTGTCTGTGGTTTGTGATGTGGCAG	1766
Db	1561	GCTGACAAAGTTCTGTGACCAAGCTGTAAAGCTTTATCTCTGCGGTTTGTGATGTGGTGAC	1620
Qy	1767	TGTGGGCAAGATCATTTTTCATGAATTTGTATAAAGTGATCCTTCTCCCAACACAGACTCAC	1826
Db	1621	TGTGGCAAGATCATTTTTCATGAATTTGTATAAAGTAAACACTTCTCCCAACACAGACTCAC	1680
Qy	1827	TATATTATTCCAAAAGGTGAATGCCCTGCCTTATTTTCAGCTTTTCGAGAAGTAGGCCAAAAGA	1886
Db	1681	TATGTTGTCCTCCAAAGGTGAATACCTGTCTTATTTTCAGCTTTTCGAAAATAGCCAGAAA	1740
Qy	1887	GGAGTTGAAGTGCCCTATAGTGACAAATCAATTAATTTCGACATGCTTCTATTTCGCAACAAG	1946
Db	1741	AGAAATTTGAAGGGACCTACAGCGACAACCCCATCATCCGCCACGCTCCATTGTGAAAACAAG	1800
Qy	1947	TGAAAACCACTCACCTCATAAATGCACAGTGGAAATGAATGCCACCAACAATATACATTTTAAAT	2006
Db	1801	TGAAAACCACTCACCTGATTAATGCCGGGGGGATGAACGCCACCAACGATCTATTTTAAC	1860
Qy	2007	CTCACGTTTCAAATAACAAACGATGAAGAGTTCAAATATGCAGATAACAGTGGAGGTGGAC	2066
Db	1861	CTCACTTTCAAACCGCAACGACGAAGAGTTCAAGATCCAGATAGCAGTAGAGGTGGAC	1920
Qy	2067	ACAAGGAGGACCAAACTGAATTTCTACGGCCGACAGAGGTTACGAAATTTAGTTTAGT	2126
Db	1921	ACGAGGAGGCGCCCAAACTGAATTTCTAACCCAGAGGCCCTATGAAGTTTGGTTAGC	1980
Qy	2127	CCCATAACTTCTTCCAGAGCGGGAATTCCTTTTTCAGGATATTCCTCCAAAGAAAAACGC	2186
Db	1981	CCAGTGACACTCTTCTCAGGCTGACGTCCTTTTGAAGATGTCCCAAGAGAAAAACGC	2040
Qy	2187	TTCCGGAAGTTTAAAGACATGATGTTTAATCTAAACAGGAGACCCACGGAAGAGGTGAAA	2246
Db	2041	TTCCCCAAGATCAGGAGACATGATGTAATATGCAACAGGGGAGATTTCGAAGAGGAGGTGAAA	2100
Qy	2247	ATTCCCTGTGTAATATTTTCACCTCTTCCAAAAGAGCCCGCAGTTTGAGTCTCAATACCTTG	2306
Db	2101	ATCCCCGGGTAAATATTTTCACCTCTTCCCAAGAGGCCCGAGTGAGGCTGAGCAACTTG	2160
Qy	2307	GATTTCGAAGTGAACATGGAGACATCACTTTTGAAGGATACAAATTTGTCCAAGTCAGCC	2366
Db	2161	GATTTGCAACTGAACGTTGAGACATCACTCTGAAAAGGATATAACTTTGTCCAAGTCAGCC	2220
Qy	2367	TTGCTGAGATCATTTCTGATGAATCAGACGATCTTAAATATAAABATCAAGCTATAATA	2426
Db	2221	CTGCTAGGTCTTTCTCTGGGGAATTCATATAGATCTATAAATAAAACCTCAAGCT---AGG	2277
Qy	2427	ACAGATGAAACAAATGACAGTTTGTGGTCTCCACAGGAAAAACAGGTTCTATAAAGCATC	2486
Db	2278	ACCGATGAAACAAAGGCAACCTGGAGTCTCCACAGAAAAACCTTCTCACAGACG----	2333
Qy	2487	TTGCCAAACAGCTTAGGAGTGTCGAAGATTCGACAGGTTTGACTTTTCTCTGAGGTGAGT	2546
Db	2334	--TCCACATGGCTTTGCTGGTGAACACAGATCAGAGAGTGGACTGCCCCAGCAGAGACA	2391
Qy	2547	GTAAAAGTGAATGGTCATGACACCGAGGTCAGAATCCACCCCTGGACTTTGAGACCAACAGCA	2606

—

Db 2392 GTGACCGTGAAGGCGGTGACCAACGCTTTGATCCACCCCGGTTGGAGACCAATGCA 2451
Qy 2607 AGATTAGAGTGGAACCTCACACCAAAAAACCATAGCGGAATGTGACAAAAAGAAAG 2666
Db 2452 AGATT-----GGCCCGCCTACACTAGCGGTGACTGTGTCACAAAGAAC 2496
Qy 2667 CCCCACCTCTGATTGTTCCACTGGAAGGCGAGATGACAAAGAAAAAGAAATCACAGGG 2726
Db 2497 CTTTCACCGCTGATCGTTCCCGGAGAACCACTTGCACAAAGAGAGGAGGTGACAGG 2556
Qy 2727 AAAGAAAAAGAGAACAGTAGAATGAGGAAAAATGCTGAAAAATCACATGAGCGTTACTGAA 2786
Db 2557 GCAGAA-----GGCAATGCTGTACCTGTAAAGGAG 2586
Qy 2787 GTGTTACTTGGAGAAAGCTGCAGCAATTACACAGATAGTTACTTGGGCTTTTGGCCATGG 2846
Db 2587 TTAGTGCCTGGCAGACGGTTGCAGC-----AGAAATATCCAGGCTTTTGGCCCTGG 2637
Qy 2847 GAGAAAAAAGATATTTCGAAGATCTTCTGCAGCAAGAGAGTCAATTGAAGACCAATTTG 2906
Db 2638 GAGAAAAAAGATATTTCGAAGATCTTCTGCAGCAAGAGAGTCAATTGAAGACCAATTTG 2697
Qy 2907 GCATACTTCACTGATAGCAAAAATCTGGGAGGCACTAAAGATACATTTGCAGATTC 2966
Db 2698 GCGTACTTTACAGACCGCAACATACCGGGAGGCACTTAAAGATACATTTGCAGATTC 2757
Qy 2967 CTCAGATATGAAAAATTTCTAAATAGCAAGTTTGGATTACATCGCGGAAAGTCCCT 3026
Db 2758 CTCGATACGTCATTAATTTCTCAACAGAGTTTGGATTCAATCCAGGAAAGTCCCT 2817
Qy 3027 GCTCATGCTTCATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCT 3086
Db 2818 GCACACATGCGGCACATGATGACAGGATCGTTATGCAAGAACTCCAAGATATGTTCCCT 2877
Qy 3087 GAAGAAATTTGACAGACGTCATTTACAAAGTGGCCCATCTGAGGATATGSCAGTTTGC 3146
Db 2878 GAAGAAATTTGACAGACGTCATTTACAAAGTGGTGTCTCTGAGGACATGCGAGTTTGC 2937
Qy 3147 TTCTCTTATTTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTT 3206
Db 2938 TTCTCTCTACTTTTATCTCATGATGATGAGTGCAGTTCAGCCCTCAATATTTCCCAAGTCTTT 2997
Qy 3207 GATGAAGTTGATACAGATCAATCTGGTGTCTGTGTGACAGAGAAATCCGACACTGGCT 3266
Db 2998 CATGAAGTAGACACAGACCAATCTGGTGTCTGTGTGATAGGAAATCCGACACTGGCC 3057
Qy 3267 ACCGAAATTCAGAACTCCGTTAAAGTTTGAGGATTTGACAGGCTCTGGAACACATGCTA 3326
Db 3058 ACGGAAATTCAGACCTACTTTAAGCTTGACAGATTTGACAGGTTTGGAAACATGTTA 3117
Qy 3327 ATAAATTCCTGTAAGATGCTCTGCTGATATACAGGAGTAAATATATTTCCACCACT 3386
Db 3118 ATAAATTCCTGTAAGATGCTCTGCTGATATACAGGAGTAAATATATTTCCACCACT 3177
Qy 3387 CAGGAATCTCTACTATGATCCAACTGCGCCCGGTGCTGATGATGATGATGATGATGATG 3446
Db 3178 CAGGAGACATACAGACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3237
Qy 3447 AAACAGTAAGTGAACAAATCCAAAGCATATTAAGGACAAAAACAAATATAGGTTTGA 3506
Db 3238 AAGCCAGTAAGTGAACAAATCCAAAGCATATTAAGGACAAAAACAAATATAGGTTTGA 3297
Qy 3507 ATCATGGGAGAGAGAAATCGCTTTTAAATGATGCTGATGATGATGATGATGATGATG 3566
Db 3298 ATCATGGGAGAGAGAAATCGCTTTTAAATGATGCTGATGATGATGATGATGATGATG 3357
Qy 3567 GCGCAGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3626
Db 3358 GGTCAAGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3417
Qy 3627 GACACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3686

Db 3418 GACCACCAATAAGATGCCCGGACAGTGAAGGCTGTCTCTCAGGCACTTCTATGAGTCC 3477
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAACCGTTTCTCTTCAT 3746
Db 3478 ATGTTCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAACCGTTTCTCTGCAC 3537
Qy 3747 ATGCTATGACTGCAGGAATGAGGGCTTATCGAGACAAATTTGAAGCTTTTGGACCAATGT 3806
Db 3538 ATGCTATGACTGCAGGAATGAGGGCTTATCGAGACAAATTTGAAGCTTTTGGACCAATGT 3597
Qy 3807 GTACTAGCAACATTTGATTTACTATTTCTCATTTTCTCATTTTCTCATTTTCTCATTTTCT 3866
Db 3598 GTACTAGCAACATTTGATTTACTATTTCTCATTTTCTCATTTTCTCATTTTCTCATTTTCT 3657
Qy 3867 CTTAAGCGGAAGATTTTCCAGAAAGGAGGATACACAAAGAGCTAGTCCCAATCGAATC 3926
Db 3658 CTGAAGCGGAAGATTTTCCAGAGAGGAGGATACACAAAGAGCTAGTCCAGACCGAATC 3717
Qy 3927 AGAGTATAGAGATCTTTCATTTGAAAAACATCTACTCTCAGCATTTTCTAGCAGATTTTAAA 3986
Db 3718 AGGGTGTAGAGATCTTTCATTTGAAAAAGTCACTTACTCTTAGCATCT- GTGAAATCTCCCT 3776
Qy 3987 ACTCAGCTTCACAGAGATGCTTCTGTGATGTGATGCTTA- GCAGTTTGGCCGGAAGAAG 4045
Db 3777 CTTGACACACAGCGGAGTCTCTGTGATGTGGACAGAGGAGCTCTGGGAGAGAAG 3836
Qy 4046 AAAATATCAGTACCATCTGTTTGTGGCATGAATATAGCCCACTGACTAGGAATTTAT 4105
Db 3837 GACATCGTGCAGACGGGTTCTTCTGCAATGGAAGAGAGAGGCCACTGACCTGGAATTTAT 3896
Qy 4106 TAACCAACCCACTGAAACTGTGTGTGAGAGCTCTGAACTGATTTTACTTTTAAAGA 4165
Db 3897 ----CAGCACTATAGAACCTGTGCAATAGC- ----TTGTACAGCTTGTACTTTTAAAGG 3948
Qy 4166 ATTTGCTCATGACCTGTCTCATCTTTTATAAAAGGCTCACTGACAGAGACAGCTGT 4225
Db 3949 ATTTGCGGAGACCTGTCTGGCTTGTGACAAACCTCCCTGACA- ----AGCTGCT 4000
Qy 4226 AATTTCCACACATCAATGACAGTAACTTTATAGGAGAGGCTATGCCAGTGGGA 4285
Db 4001 GGTTCCTTCCCGAGTACTGCAGACT-----GAGAAACCACTCCATCTTGAAA 4049
Qy 4286 GTGATTGTCTAAGAGGCTCCAGTCTTGTGATTCGAAAGCTTTTGTAAAGTTTGTGCACTT 4345
Db 4050 GCAAGTGGGAGGGGCCAGTCTTGTGATTCGAAAGCTTTCCAGCATAAATTTCTGGCTT 4109
Qy 4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTACTAAGTTAACT 4388
Db 4110 GTCTCTCTTTTGTATCCATTTTCCATTTTAAAAAACAATTAAGTGGCTACTAAGTT 4169
Qy 4389 AGTTATCTTGTCTCTGAGTATAACGAAATGGGATGTCTAAACCTATTTTATATAGATCT 4448
Db 4170 AGTCATTTCTCACTTTTCAAAATTAACAAATCAGGATGTCAAAAC- -ATTGTATAGATCT 4227
Qy 4449 ATTTAAATTAATGACCAATATCACCTCTTATTTGACAAATACCTAAATTAAGAGTTTATTA 4508
Db 4228 ATTTAAATTAATAGAACGATTACTTCT- -TTAGCTATCTAAATTAATTTTATTTATTA 4285
Qy 4509 ATATTTAAGACTGTAAATGGTCTTAAACCACT- -AATCTCTGAGAGCTCAATCATTTGA 4565
Db 4286 ACAGTCAAG-----TGGTCTTGAACCGCTTAACTACTTCCCAAGAAATCTGTGACCTGTGTC 4336
Qy 4566 CATCTGAAATGCTTGTGTAATTTATGACTTTCAGCCCTCAAGAAATGCTATGATTTACGTGC 4625
Db 4337 CGTTGAAAGTGTCTTGGAGCTTGTAACTCAATTTCCCAAGAAATCTGTGACCTGTGTC 4396
Qy 4626 AGGTCTAATTTCAACAGGCTAGAT-----TAGTACTCTTACAGATGATTAATGT 4678
Db 4397 GGGCTGTGTCGAAAGGGCTAGTGTGACGTAGCTGTGCTCACCAGATGTAATTTATGT 4456
Qy 4679 TTTGAAATGATACATATTCAAAACAGAAAGTGTCTCATTTTGAAGAAATGAGTGTGCTGATGG 4738
Db 4457 CGTGGAAATGTACATA--CAGACAAAGTGTCTCACTTCAGAAATGAGTGTGCTGATGG 4514

Qy 4739 CACTGGCACATTACAGTG 4756
|||
Db 4515 CACCAGCGAGTGATGGTG 4532

Search completed: November 22, 2004, 06:14:41
Job time : 22750 secs

2004 Nov 28 09:38:36

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 18:18:51 ; Search time 2385 Seconds
(without alignments)
12319.082 Million cell updates/sec

Title: US-10-023-888-3
Perfect score: 5597
Sequence: 1 cggagcagagggcgctccg.....aaaaagttaattttgaaa 5597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5597	100.0	5597	4 ACC81001	Acc81001 Human Glc
2	5597	100.0	5597	10 ADD27811	Add27811 Human Glc
3	5597	100.0	5597	10 RAD62491	Rad62491 Human Glc
4	5597	100.0	5597	10 RAD62650	Rad62650 Human Glc
5	4326	77.3	4369	6 AD41199	Ad41199 Human nuc
6	4057	72.5	4277	6 AAD41203	Aad41203 Human nuc
7	3761	67.2	3783	4 ACC81010	Acc81010 Human Glc
8	3485.4	62.3	3621	4 ACC81011	Acc81011 Human Glc
9	3465.6	61.9	3600	10 ADD27809	Add27809 Soluble h
10	3465.6	61.9	3600	10 RAD62490	Rad62490 N-acetylgl
11	3465.6	61.9	3600	10 RAD62649	Rad62649 N-acetylgl
12	2820.8	50.4	5229	4 ACC81007	Acc81007 Murine Gl
13	2820.8	50.4	5229	10 ADD27816	Add27816 GlcNAc-ph
14	2820.8	50.4	5229	10 RAD62493	Rad62493 Mouse cDN
15	2820.8	50.4	5229	10 AAD62652	Aad62652 Mouse DNA
16	2613	45.7	8810	10 ADG39793	Adg39793 Human cDN
17	2613	45.7	8811	8 ABX10235	Abx10235 Human cDN
18	2587.4	46.2	2603	10 ADAS3185	Ada53185 Human cod
19	2372.2	42.4	2428	6 AAL50831	Aal50831 Human can
20	2018.2	36.1	2076	4 AAH14758	Aah14758 Human cDN
21	1716.4	30.7	16800	4 AAK82065	Aak82065 Human imm

C	22	1716.4	30.7	16805	4	AAK82067	Aak82067 Human imm
C	23	1716.4	30.7	16806	4	AAK82066	Aak82066 Human imm
C	24	1105.2	19.7	1194	12	ACH89502	Ach89502 Human gen
C	25	1104.8	19.7	1962	4	AAI12977	Aai12977 Probe #29
C	26	1104.8	19.7	1962	4	AAI34336	Aai34336 Probe #30
C	27	1104.8	19.7	1962	4	ABA44229	Aba44229 Human bre
C	28	1104.8	19.7	1962	4	AAK02966	Aak02966 Human bra
C	29	1104.8	19.7	1962	5	AAI02895	Aai02895 Probe #28
C	30	1104.8	19.7	1962	5	ABS02920	Abs02920 Human gen
C	31	1103.4	19.7	1125	4	AAI22215	Aai22215 Probe #12
C	32	1103.4	19.7	1125	4	AAI47512	Aai47512 Probe #16
C	33	1103.4	19.7	1125	4	ABA49383	Aba49383 Human bra
C	34	1103.4	19.7	1125	4	AAK15732	Aak15732 Human bra
C	35	1103.4	19.7	1125	5	AAI07914	Aai07914 Probe #79
C	36	1103.4	19.7	1125	6	ABS15475	Abs15475 Human gen
C	37	724.8	12.9	728	10	ADD26805	Add26805 Human adi
C	38	501	9.0	526	2	AAH07289	Aah07289 Human cDN
C	39	486.2	8.7	830	8	ABX34834	Abx34834 EST clone
C	40	486.2	8.7	830	8	ACH38230	Ach38230 Human mdd
C	41	415.4	7.4	465	9	ACH38230	Ach38230 Human end
C	42	398.4	7.1	537	3	AAA43989	Aaa43989 Mouse sec
C	43	377.2	6.7	459	9	ACH40463	Ach40463 Human foe
C	44	356.2	6.4	392	4	AAH12279	Aah12279 Human cDN
C	45	353.4	6.3	773	10	ADI21775	Adi21775 Novel hum

ALIGNMENTS

RESULT 1
ACCS81001
ID ACC81001 standard; cDNA; 5597 BP.
XX
AC ACC81001;
XX
DT 01-AUG-2003 (first entry)
XX
DE Human GlcNAc-phosphotransferase alpha/beta-subunit cDNA.
XX
KW Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW lysosomal storage disease; enzyme; alpha-subunit; gene; ss; beta-subunit.
XX Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 185..2948
FT /*tag= a
FT /product= "GlcNAc-phosphotransferase alpha subunit"
FT /note= "No stop codon given"
FT CDS 2949..3935
FT /*tag= b
FT /product= "GlcNAc-phosphotransferase beta subunit"
FT /note= "No start codon"
XX
US6537785-B1.
25-MAR-2003.
10-AUG-2000; 2000US-00636077.
14-SEP-1999; 99US-0153831P.
(GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.
Canfield WM;
WPI; 2001-290356/30.
P-FSDB; ABR61376, ABR61377.
Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-

QY 1861 TCAGCTTTGCGAGTACGCAAGAGAGAGTGGAGTGCCTATAGTACCAATCAATAA 1920
 Db 1861 TCAGCTTTGCGAGTACGCAAGAGAGAGTGGAGTGCCTATAGTACCAATCAATAA 1920
 QY 1921 TTCGACATGCTTTCTATTTGCCAACAGTGGAAACCAATCCACTCATATGACAGTGGAA 1980
 Db 1921 TTCGACATGCTTTCTATTTGCCAACAGTGGAAACCAATCCACTCATATGACAGTGGAA 1980
 QY 1981 TGAATGCCACCAATACATTTTAAATCTCAGTTTCAAAATACAAACGATGAAGAGTTCA 2040
 Db 1981 TGAATGCCACCAATACATTTTAAATCTCAGTTTCAAAATACAAACGATGAAGAGTTCA 2040
 QY 2041 AAATCCAGATACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAAATCTACGGGCC 2100
 Db 2041 AAATCCAGATACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAAATCTACGGGCC 2100
 QY 2101 AGAAGGTTAGCAAAATTTAGTATGTCCTATTAACACTTTCTCCAGAGCGGAAATCTTT 2160
 Db 2101 AGAAGGTTAGCAAAATTTAGTATGTCCTATTAACACTTTCTCCAGAGCGGAAATCTTT 2160
 QY 2161 TTGAGGATATCCCAAGAAACCGTTCCCGAGGTTTAAAGACATGATGTTAACTCAA 2220
 Db 2161 TTGAGGATATCCCAAGAAACCGTTCCCGAGGTTTAAAGACATGATGTTAACTCAA 2220
 QY 2221 CAAGGAGAGCCAGGAAGAGTGAATAATCCCTGGTAAATTTTCACTCTCTCCAAAG 2280
 Db 2221 CAAGGAGAGCCAGGAAGAGTGAATAATCCCTGGTAAATTTTCACTCTCTCCAAAG 2280
 QY 2281 AGCCCGATGAGTCTCAATACCTTGGATTTGCAACTGGACATGAGACATCTTTGA 2340
 Db 2281 AGCCCGATGAGTCTCAATACCTTGGATTTGCAACTGGACATGAGACATCTTTGA 2340
 QY 2341 AAGGATACAAATTTGTCAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 2400
 Db 2341 AAGGATACAAATTTGTCAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 2400
 QY 2401 CTAAATATAAATCAAGTATTAATAACAGATGAACAAATGAAGTGGTGGCTCCAC 2460
 Db 2401 CTAAATATAAATCAAGTATTAATAACAGATGAACAAATGAAGTGGTGGCTCCAC 2460
 QY 2461 AGGAAACAGAGTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGATGTC 2520
 Db 2461 AGGAAACAGAGTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGATGTC 2520
 QY 2521 AGAGTTGACTTTTCTGAGTGGTGAAGTGAATGGTCAATGACAGGCTCAGAAATC 2580
 Db 2521 AGAGTTGACTTTTCTGAGTGGTGAAGTGAATGGTCAATGACAGGCTCAGAAATC 2580
 QY 2581 CACCCCTGAGCTTGAGACCAAGCAAGATTTAGAGTGGAACTCAGACCCCAAAACCA 2640
 Db 2581 CACCCCTGAGCTTGAGACCAAGCAAGATTTAGAGTGGAACTCAGACCCCAAAACCA 2640
 QY 2641 TAGGCGAAATGTGACAAAGAAAGCCCAATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700
 Db 2641 TAGGCGAAATGTGACAAAGAAAGCCCAATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700
 QY 2701 TGACAAAGAAAGAAATCAAGGAAAGAAAGAGAGACAGTGAAGTGGAGGAAATG 2760
 Db 2701 TGACAAAGAAAGAAATCAAGGAAAGAAAGAGAGACAGTGAAGTGGAGGAAATG 2760
 QY 2761 CTGAAATCAGATAGCGGTTACTGAGTGTGTTGGAAGAGCTGAGAGTACACAG 2820
 Db 2761 CTGAAATCAGATAGCGGTTACTGAGTGTGTTGGAAGAGCTGAGAGTACACAG 2820
 QY 2821 ATAGTTACTTGGCTTTTGGCATGGGAAAGAAAGTATTTCCAGATCTTCTCGAG 2880
 Db 2821 ATAGTTACTTGGCTTTTGGCATGGGAAAGAAAGTATTTCCAGATCTTCTCGAG 2880
 QY 2881 AAGAGAGTCAATGAGACACATTTGGCATCTTCACTGATGACAAATCTGGAGGC 2940
 Db 2881 AAGAGAGTCAATGAGACACATTTGGCATCTTCACTGATGACAAATCTGGAGGC 2940

QY 2941 AACTAAAGATACATTTGCGAGTTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
 Db 2941 AACTAAAGATACATTTGCGAGTTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
 QY 3001 TTGGAATTCATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
 Db 3001 TTGGAATTCATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
 QY 3061 TGCAGAACTGCAAGATATGTTCCCTGGAAGATTTGCAAGAGCTCATTTCAAAAGTGC 3120
 Db 3061 TGCAGAACTGCAAGATATGTTCCCTGGAAGATTTGCAAGAGCTCATTTCAAAAGTGC 3120
 QY 3121 GCCATTCTGAGGATATGCGAGTTCCTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180
 Db 3121 GCCATTCTGAGGATATGCGAGTTCCTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180
 QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTCATACAGATCAATCTGGTGTCTGT 3240
 Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTCATACAGATCAATCTGGTGTCTGT 3240
 QY 3241 CTGACAGAGAAATCCGAAACACTGCTTACCGAATTCACGAACTGCCGTTAAGTTTGCAGG 3300
 Db 3241 CTGACAGAGAAATCCGAAACACTGCTTACCGAATTCACGAACTGCCGTTAAGTTTGCAGG 3300
 QY 3301 ATTGTACAGTCTGGAAACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
 Db 3301 ATTGTACAGTCTGGAAACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
 QY 3361 CGGAGCTTAATATATTTCCACCAACTCAGAAATCTTACTATGATCCCAACTGCCACCGG 3420
 Db 3361 CGGAGCTTAATATATTTCCACCAACTCAGAAATCTTACTATGATCCCAACTGCCACCGG 3420
 QY 3421 TCCTAAAGTCTAGTAACAAACTGTAAACCACTTAACCACTGACAAATCCCAAGCATATA 3480
 Db 3421 TCCTAAAGTCTAGTAACAAACTGTAAACCACTTAACCACTGACAAATCCCAAGCATATA 3480
 QY 3481 AGGACAAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGA 3540
 Db 3481 AGGACAAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGA 3540
 QY 3541 TTGCTACCAACGTTTCTCATGTTGGCAGTGGATGACATGAAGAAACCCCTAGGA 3600
 Db 3541 TTGCTACCAACGTTTCTCATGTTGGCAGTGGATGACATGAAGAAACCCCTAGGA 3600
 QY 3601 AGTTGTTGCTGGAATGACAACTGACCAATCATAAAGTGTCTGACAGTGAAG 3660
 Db 3601 AGTTGTTGCTGGAATGACAACTGACCAATCATAAAGTGTCTGACAGTGAAG 3660
 QY 3661 CTGTTCTCAGGACTTCTATGATCCATGTTCCCACTCCTTCCCAATTTGAACCTGCCAA 3720
 Db 3661 CTGTTCTCAGGACTTCTATGATCCATGTTCCCACTCCTTCCCAATTTGAACCTGCCAA 3720
 QY 3721 GAGAGTATCGAAACCGTTTCTCATGATGCTGAGGAAATGGAGGCTTATCGAG 3780
 Db 3721 GAGAGTATCGAAACCGTTTCTCATGATGCTGAGGAAATGGAGGCTTATCGAG 3780
 QY 3781 ACAATTTGAAGTTCGAGCCCATGTTACTAGCAACATTTGTTTACTATATCT 3840
 Db 3781 ACAATTTGAAGTTCGAGCCCATGTTACTAGCAACATTTGTTTACTATATCT 3840
 QY 3841 CATTTTTCGAGAGTAAATTTGCACTTAAAGCGAAGATATTTCCAGAGGAGGATAC 3900
 Db 3841 CATTTTTCGAGAGTAAATTTGCACTTAAAGCGAAGATATTTCCAGAGGAGGATAC 3900
 QY 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCATCTA 3960
 Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCATCTA 3960
 QY 3961 CCTCAGATTTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTAT 4020
 Db 3961 CCTCAGATTTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTAT 4020
 QY 4021 GCTTAGCAGTTTGGCCCCGAAAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080


```
Db 4021 GCTTAGCAGTTTGGCCCGAAGAAATATCCAGTACCAGTCTGTTTGTGGCAGAA 4080
QY 4081 TATAGCCCACTGACTAGGAATATTTTAAACCAACCACTGAAACCTGTGTGTGAGCAGC 4140
Db 4081 TATAGCCCACTGACTAGGAATATTTTAAACCAACCACTGAAACCTGTGTGTGAGCAGC 4140
QY 4141 TCTGAACCTGATTTTACCTTTTAAAGATTTGCTATGAGCCTGTCATCTTTTATATAAAA 4200
Db 4141 TCTGAACCTGATTTTACCTTTTAAAGATTTGCTATGAGCCTGTCATCTTTTATATAAAA 4200
QY 4201 GGCTCACTGACAAGACAGCTGTTTAAATTTCCCAAGCAATCATTTGACAGCTAACTTTAT 4260
Db 4201 GGCTCACTGACAAGACAGCTGTTTAAATTTCCCAAGCAATCATTTGACAGCTAACTTTAT 4260
QY 4261 TAGGAGAGCCTATGCCAGCTGGAGTGATTTGCTTAAAGAGGCTCCAGTCTTTGCAATCCAA 4320
Db 4261 TAGGAGAGCCTATGCCAGCTGGAGTGATTTGCTTAAAGAGGCTCCAGTCTTTGCAATCCAA 4320
QY 4321 AGCCTTTTGTCTAAAGTTTGCACCTTTTATTTTTCATTTCCCATTTTAAAGTACTACTA 4380
Db 4321 AGCCTTTTGTCTAAAGTTTGCACCTTTTATTTTTCATTTCCCATTTTAAAGTACTACTA 4380
QY 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATGGGATGCTTAAACCTATTTTAA 4440
Db 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATGGGATGCTTAAACCTATTTTAA 4440
QY 4441 TAGATGTTTATTAATAATGACGAATATACCTCTTATTTGACAAATACCTAAATATGAG 4500
Db 4441 TAGATGTTTATTAATAATGACGAATATACCTCTTATTTGACAAATACCTAAATATGAG 4500
QY 4501 TTTTATTAATATTTAAAGACTGTAATGCTCTTAAACCACTAACTACTGAGAGCTCAATG 4560
Db 4501 TTTTATTAATATTTAAAGACTGTAATGCTCTTAAACCACTAACTACTGAGAGCTCAATG 4560
QY 4561 ATTGACATCTGAAATGCTTTGTAATATTTAGCTTCAAGCCCTTAAGAAATGCTATGATTTCA 4620
Db 4561 ATTGACATCTGAAATGCTTTGTAATATTTAGCTTCAAGCCCTTAAGAAATGCTATGATTTCA 4620
QY 4621 CGTCAGCTCTAAATTTCAACAGGCTAGAGTTAGTACTTACAGAGTGAATATGTTT 4680
Db 4621 CGTCAGCTCTAAATTTCAACAGGCTAGAGTTAGTACTTACAGAGTGAATATGTTT 4680
QY 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGAAATGAGTACTGATGGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGAAATGAGTACTGATGGCA 4740
QY 4741 CTGGCAGATTTACAGTGGTCTCTTTTAAATCTCATTTGGTATATTCAGAGTACTTCT 4800
Db 4741 CTGGCAGATTTACAGTGGTCTCTTTTAAATCTCATTTGGTATATTCAGAGTACTTCT 4800
QY 4801 CTCAGTTGGTTTTGATAGACAGAGCCAGCAAACTTTCTTTGTAAGGCTGGTTAGT 4860
Db 4801 CTCAGTTGGTTTTGATAGACAGAGCCAGCAAACTTTCTTTGTAAGGCTGGTTAGT 4860
QY 4861 AAATTTTGGAGCCAGCTGCTTTGTCTATACATTTCTTCTGCTGCTTTTGTAGTTGT 4920
Db 4861 AAATTTTGGAGCCAGCTGCTTTGTCTATACATTTCTTCTGCTGCTTTTGTAGTTGT 4920
QY 4921 TTTTCTTCAAAACAACTCTTAAATGTAAACCACTGTTTAGCTTCCAGCTGTACAAA 4980
Db 4921 TTTTCTTCAAAACAACTCTTAAATGTAAACCACTGTTTAGCTTCCAGCTGTACAAA 4980
QY 4981 ACTGCCCAACAGCAGATGTGACCTCTCAGGCCATCATTTGCCAATCATCTGAGAAATATTT 5040
Db 4981 ACTGCCCAACAGCAGATGTGACCTCTCAGGCCATCATTTGCCAATCATCTGAGAAATATTT 5040
QY 5041 TTGTTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTGCCAGGCTGGAG 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTGCCAGGCTGGAG 5100
QY 5101 TGCAGTGGCGCAATCTCAGCTCATCTGCAAACTCCGCTCCCGGTTCAACAGTTCTGTC 5160
```

```
5101 TGCAGTGGCGCAATCTCAGCTCATCTGCAAACTCCGCTCCCGGTTCAACAGTTCTGTC 5160
QY 5161 TCAGCCTTCTGAGTAGCTGGGACTACAGGTGCATGCCACCAACCTCTAATTTTGTGTA 5220
Db 5161 TCAGCCTTCTGAGTAGCTGGGACTACAGGTGCATGCCACCAACCTCTAATTTTGTGTA 5220
QY 5221 TTTTGTAGTAGACACGGGGTTCCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280
Db 5221 TTTTGTAGTAGACACGGGGTTCCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280
QY 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTCTGAGATTACAGGCATAAGCCAGTGCAC 5340
Db 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTCTGAGATTACAGGCATAAGCCAGTGCAC 5340
QY 5341 CCAGCCGGAATAGTATTTTATGTATGTTTAAACCTTGGGCTCTAGCCATATTTATG 5400
Db 5341 CCAGCCGGAATAGTATTTTATGTATGTTTAAACCTTGGGCTCTAGCCATATTTATG 5400
QY 5401 TCATAATAACAATGATTTGTGAAGACAGATTCCATGAGTAACTCTGACAGGTATTTTATG 5460
Db 5401 TCATAATAACAATGATTTGTGAAGACAGATTCCATGAGTAACTCTGACAGGTATTTTATG 5460
QY 5461 ATCATGATCTCAACAATATTTCTCCCAATGSCATACATCTTTTGTACAAAGACTTGAA 5520
Db 5461 ATCATGATCTCAACAATATTTCTCCCAATGSCATACATCTTTTGTACAAAGACTTGAA 5520
QY 5521 ATGTAATAACTGTTGTTGCTGTGAAGAGTGTCTATTTCAAACCTGAAATCTCATAAA 5580
Db 5521 ATGTAATAACTGTTGTTGCTGTGAAGAGTGTCTATTTCAAACCTGAAATCTCATAAA 5580
QY 5581 AAGTTAAATTTTGA AAA 5597
Db 5581 AAGTTAAATTTTGA AAA 5597

RESULT 2
ADD27811
ID ADD27811 standard; cDNA; 5597 BP.
XX AC ADD27811;
XX AC ADD27811;
XX DT 15-JAN-2004 (first entry)
XX DE Human GlcNAc-phosphotransferase alpha/beta subunit precursor cDNA.
XX DE
XX KW human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
XX KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ss;
XX KW gene.
XX OS Homo sapiens.
XX FH Key
XX FT CDS
XX FT 165..2948
XX FT /*tag= a
XX FT /note= "Specifically claimed in claim 5"
XX FT /product= "alpha subunit"
XX FT 2949..3932
XX FT /*tag= b
XX FT /note= "Specifically claimed in claim 6"
XX FT /product= "beta subunit"
XX PN US2003119088-A1.
XX PD 26-JUN-2003.
XX PF 21-DEC-2001; 2001US-00023888.
XX PR 21-DEC-2001; 2001US-00023888.
XX PA (NOVA-) NOVAZYME PHARM INC.
XX PI Canfield W, Kudo M;
```

DR WPI; 2003-801323/75.
 DR P-PSDB; ADD27812, ADD27813.
 XX Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 XX protein.
 XX
 PS Disclosure; SEQ ID NO 3; 55pp; English.
 XX
 CC The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. the method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the human GlcNAc-
 CC phosphotransferase alpha/beta subunit precursor cDNA.
 XX
 SQ Sequence 5597 BP; 1678 A; 1151 C; 1219 G; 1549 T; 0 U; 0 Other;
 Query Match 100.0%; Score 5597; DB 10; Length 5597;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCGGAGGCTGTGACC 60
 DB 1 CGGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCGGAGGCTGTGACC 60
 QY 61 TCGCGCGCGCGCCCGACCGGGCCCTGAATGCGCGCTCGCTGAGCGCGCGCGCGCGC 120
 DB 61 TCGCGCGCGCGCGCGACCGGGCCCTGAATGCGCGCTCGCTGAGCGCGCGCGCGCGC 120
 QY 121 GCGGCTCAGGCTCCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180
 DB 121 GCGGCTCAGGCTCCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180
 QY 181 TGCAGAGACAACTATACCTGCTGCTCCAGAGTATGGGCTCTACGCTGTCTTCTGG 240
 DB 181 TGCAGAGACAACTATACCTGCTGCTCCAGAGTATGGGCTCTACGCTGTCTTCTGG 240
 QY 241 GCGTGTGTACCATGCTCCGCTTCAGTTCGAGAGGTGTTCTGGAATGGAGCC 300
 DB 241 GCGTGTGTACCATGCTCCGCTTCAGTTCGAGAGGTGTTCTGGAATGGAGCC 300
 QY 301 GAGATCAATACCATGTTGTTGTTGATTCCTATAGAGACAAATTCCTGGAAGTCCCTTC 360
 DB 301 GAGATCAATACCATGTTGTTGTTGATTCCTATAGAGACAAATTCCTGGAAGTCCCTTC 360
 QY 361 AGAATCGGCTTGTCTGCCATGCGCATGACGTTGTTTACCTGGGTGAATGGCACAG 420
 DB 361 AGAATCGGCTTGTCTGCCATGCGCATGACGTTGTTTACCTGGGTGAATGGCACAG 420
 QY 421 ATCTTTGAATCTGAAGGAACCTACAGCAGGTCAGAGNACAGATGGAGGAGGAGCAAG 480
 DB 421 ATCTTTGAATCTGAAGGAACCTACAGCAGGTCAGAGNACAGATGGAGGAGGAGCAAG 480
 QY 481 CAATGAGAGAAATCCTTGGGAAAAACAAACGGAACTTAAAGAGAGTGAAGAGCAGT 540
 DB 481 CAATGAGAGAAATCCTTGGGAAAAACAAACGGAACTTAAAGAGAGTGAAGAGCAGT 540
 QY 541 TAGAGTGTGTGTAACACTGCAATTAAGTGGCAATGCTTGTACTGACCCAGCCCTGC 600
 DB 541 TAGAGTGTGTGTAACACTGCAATTAAGTGGCAATGCTTGTACTGACCCAGCCCTGC 600
 QY 601 CAGCCAAACATCACCCTGAAGGACGTCATCTCTTTATCTCTTTTCATTTCTGCCAGT 660
 DB 601 CAGCCAAACATCACCCTGAAGGACGTCATCTCTTTATCTCTTTTCATTTCTGCCAGT 660
 QY 661 ACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCACGTTGTTGTTTG 720
 DB 661 ACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCACGTTGTTGTTTG 720
 QY 721 ACAGTACTAAGGATGTGAGATGCCACCTCTGAGCTGCTTAAAGGAATAGCAGACAGA 780

DB 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTCTGACTGCTTAAAGGAAATAGCAGACAGA 780
 QY 781 CAGTATGGAGGGGTACTTTCACAAACAGATAAAGAGTCCCTGATTAAGTCTTAATGCAAG 840
 DB 781 CAGTATGGAGGGGTACTTTCACAAACAGATAAAGAGTCCCTGATTAAGTCTTAATGCAAG 840
 QY 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAAGGAAACAAATCAACTAAAAACAA 900
 DB 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAAGGAAACAAATCAACTAAAAACAA 900
 QY 901 AATTGCCAGAAATCTTCTCTTAAGTCAAACTGTTGCAGTTGATTCAGAGGCCAGTG 960
 DB 901 AATTGCCAGAAATCTTCTCTTAAGTCAAACTGTTGCAGTTGATTCAGAGGCCAGTG 960
 QY 961 TAGCGCTTCTAAACCTGAATAACCCCAAGGATTTCAAGAATTGAATAGCAAACTAAGA 1020
 DB 961 TAGCGCTTCTAAACCTGAATAACCCCAAGGATTTCAAGAATTGAATAGCAAACTAAGA 1020
 QY 1021 AGAATATGACCATGATGAGAAAGACTGACATAAGTCTCTGATATTTATTTATGGATC 1080
 DB 1021 AGAATATGACCATGATGAGAAAGACTGACATAAGTCTCTGATATTTATTTATGGATC 1080
 QY 1081 TGAGCGCATCAGCGAGTCTTAAGCAGGATCAAGACATCTCTGCCAGTCGTTTTGAAGATA 1140
 DB 1081 TGAGCGCATCAGCGAGTCTTAAGCAGGATCAAGACATCTCTGCCAGTCGTTTTGAAGATA 1140
 QY 1141 ACAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGTGCACCATGCGTTTCGGAATA 1200
 DB 1141 ACAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGTGCACCATGCGTTTCGGAATA 1200
 QY 1201 TTTTCTATGTCACCAACGGGCGATTTCCATCTCGGCTGAACTTGAACATCTCTCGAGTGA 1260
 DB 1201 TTTTCTATGTCACCAACGGGCGATTTCCATCTCGGCTGAACTTGAACATCTCTCGAGTGA 1260
 QY 1261 CAATAGTAACACACCCAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320
 DB 1261 CAATAGTAACACACCCAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320
 QY 1321 CTGCTATTGAAGTCACTCATCGCATCGAAGGCTGTCAGAGAGTATTTATTTACTATA 1380
 DB 1321 CTGCTATTGAAGTCACTCATCGCATCGAAGGCTGTCAGAGAGTATTTATTTACTATA 1380
 QY 1381 ATGATGATGTCATGTTTGGAGAGATGCTGCGCCAGATGATTTTACAGTCACTCCAAAG 1440
 DB 1381 ATGATGATGTCATGTTTGGAGAGATGCTGCGCCAGATGATTTTACAGTCACTCCAAAG 1440
 QY 1441 GCCAGAGGTTTATTTGACATGCGCTGTCGCAAACTGTCGCGAGGCTGCCAGGTTCT 1500
 DB 1441 GCCAGAGGTTTATTTGACATGCGCTGTCGCAAACTGTCGCGAGGCTGCCAGGTTCT 1500
 QY 1501 GGATTAAGGATGCTTATTTGACAGGCTTGTAAATTAATTTAGCTGCGATTTGGATGGTG 1560
 DB 1501 GGATTAAGGATGCTTATTTGACAGGCTTGTAAATTAATTTAGCTGCGATTTGGATGGTG 1560
 QY 1561 GGGATTTGCTGGAACAGTGGAGGAGTCCGCTATATTTGAGGAGTGGAGGTTCTGGGA 1620
 DB 1561 GGGATTTGCTGGAACAGTGGAGGAGTCCGCTATATTTGAGGAGTGGAGGTTCTGGGA 1620
 QY 1621 GTATTTGGAGTGGACACCCCTGCGAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680
 DB 1621 GTATTTGGAGTGGACACCCCTGCGAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680
 QY 1681 ATCAGGAGATGCGAAATTCCTGCTCGCTGATAAGTTCGTGACCAAGCATGCAATGTCT 1740
 DB 1681 ATCAGGAGATGCGAAATTCCTGCTCGCTGATAAGTTCGTGACCAAGCATGCAATGTCT 1740
 QY 1741 TGTCTGTGGGTTTGTGCTGGGACATGTTGGGCAAGATCATTTTCATGATTTGTATAAG 1800
 DB 1741 TGTCTGTGGGTTTGTGCTGGGACATGTTGGGCAAGATCATTTTCATGATTTGTATAAG 1800
 QY 1801 TGATCTTCTCCCAACCCAGACTCACTATATTTCCAAAGGTGAATGCTGCCTTATT 1860
 DB 1801 TGATCTTCTCCCAACCCAGACTCACTATATTTCCAAAGGTGAATGCTGCCTTATT 1860

QY	1851	TCAGCTTTCGAGAGTAGCCAAAGAGAGGAGTTGAAGTGCTTAGTGACAAATCCAAATAA	1920	QY	2941	AACTAAAGATACATTTGCGAGATCCCTCAGATATGTAATAATAAATTTCTAAATAGCAAGT	3000
DB	1861	TCAGCTTTCGAGAGTAGCCAAAGAGAGGAGTTGAAGTGCTTAGTGACAAATCCAAATAA	1930	DB	2941	AACTAAAGATACATTTGCGAGATCCCTCAGATATGTAATAATAAATTTCTAAATAGCAAGT	3000
QY	1921	TTGACATGCTTTCTATTGCGAAACAAAGTGGAAACCATCCACCTCATTAATGACAGTGGAA	1980	QY	3001	TTGGATTTCACATCGCGGAAAGTCCCTGCTCATGCGCTCAGATGATGACCGGATTTGTTA	3060
DB	1921	TTGACATGCTTTCTATTGCGAAACAAAGTGGAAACCATCCACCTCATTAATGACAGTGGAA	1990	DB	3001	TTGGATTTCACATCGCGGAAAGTCCCTGCTCATGCGCTCAGATGATGACCGGATTTGTTA	3060
QY	1981	TGAATGCCACCAACATACATTTTAATCTCAGCTTTCAAATATACAAACGATGAAGAGTTCA	2040	QY	3061	TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGCTCAATTCACAAAGTGC	3120
DB	1981	TGAATGCCACCAACATACATTTTAATCTCAGCTTTCAAATATACAAACGATGAAGAGTTCA	2040	DB	3061	TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGCTCAATTCACAAAGTGC	3120
QY	2041	AAATGACAGTAACAGTGGAGTGACACAAAGGAGGACCAAAACCTGAATCTACGGCC	2100	QY	3121	GCATTTCTGAGATATGAGTTGCGCTTCTCTATTATTTATATCTCATGATGCGAGTGC	3180
DB	2041	AAATGACAGTAACAGTGGAGTGACACAAAGGAGGACCAAAACCTGAATCTACGGCC	2100	DB	3121	GCATTTCTGAGATATGAGTTGCGCTTCTCTATTATTTATATCTCATGATGCGAGTGC	3180
QY	2101	AGAGGGTTACGAAATTTAGTGTCCATAACACTTTCCAGAGGCGGAAATCCCTT	2160	QY	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT	3240
DB	2101	AGAGGGTTACGAAATTTAGTGTCCATAACACTTTCCAGAGGCGGAAATCCCTT	2160	DB	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT	3240
QY	2161	TTGAGGATATTTCCAAAGAAACAGCTTCCGAAAGTTTAAGAGACATGATGTTAACTCAA	2220	QY	3241	CTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG	3300
DB	2161	TTGAGGATATTTCCAAAGAAACAGCTTCCGAAAGTTTAAGAGACATGATGTTAACTCAA	2220	DB	3241	CTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG	3300
QY	2221	CAAGGAGGCGGAGAGAGTGAATTTCCCTGTTAAATATTTCACTGCTTCCAAAG	2280	QY	3301	ATTTGACAGGCTGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360
DB	2221	CAAGGAGGCGGAGAGAGTGAATTTCCCTGTTAAATATTTCACTGCTTCCAAAG	2280	DB	3301	ATTTGACAGGCTGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360
QY	2281	ACGCCAGTTCAGTCTCAATACCTTGGATTGCAATGGAACATGGAGACATCACTTTGA	2340	QY	3361	CGCAGCTAAATAATATTTCCCAACTCAGGAATCTACTATGATCCCAACCTGCCACCGG	3420
DB	2281	ACGCCAGTTCAGTCTCAATACCTTGGATTGCAATGGAACATGGAGACATCACTTTGA	2340	DB	3361	CGCAGCTAAATAATATTTCCCAACTCAGGAATCTACTATGATCCCAACCTGCCACCGG	3420
QY	2341	AAGGATCAATTTGTCGAAGTTCAGCTTGTGAGATCATTTCTGATGAACTCAGCATG	2400	QY	3421	TCACCTAAAGTCTAGTAACCACTGTAACCAAGTAACCTGACAAATCCACAAAGCATATA	3480
DB	2341	AAGGATCAATTTGTCGAAGTTCAGCTTGTGAGATCATTTCTGATGAACTCAGCATG	2400	DB	3421	TCACCTAAAGTCTAGTAACCACTGTAACCAAGTAACCTGACAAATCCACAAAGCATATA	3480
QY	2401	CTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGAGTGTGCTCCAC	2460	QY	3481	AGGACAAAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540
DB	2401	CTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGAGTGTGCTCCAC	2460	DB	3481	AGGACAAAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540
QY	2461	AGGAAAAACAGGTTTCAAAAAGCATTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC	2520	QY	3541	TTGCTACCAAGCTTCTCATGTGTTGSCCAGTTGGATGACATAGAAAAACCTTAGGA	3600
DB	2461	AGGAAAAACAGGTTTCAAAAAGCATTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC	2520	DB	3541	TTGCTACCAAGCTTCTCATGTGTTGSCCAGTTGGATGACATAGAAAAACCTTAGGA	3600
QY	2521	AGAGGTTGACATTTCTGAGTGTGTAAGTGAATGATGATGATGATGATGATGATGATG	2580	QY	3601	AGTTTGTGCTGAAATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAG	3660
DB	2521	AGAGGTTGACATTTCTGAGTGTGTAAGTGAATGATGATGATGATGATGATGATGATG	2580	DB	3601	AGTTTGTGCTGAAATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAG	3660
QY	2581	CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCCAAAAAACCA	2640	QY	3661	CTGTTCTCAGGACCTTCTATGAACTCATGTTCCCATACCTTCCCAATTTGAACTGCCAA	3720
DB	2581	CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCCAAAAAACCA	2640	DB	3661	CTGTTCTCAGGACCTTCTATGAACTCATGTTCCCATACCTTCCCAATTTGAACTGCCAA	3720
QY	2641	TAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATGTTCCACTGGAAGCCAGA	2700	QY	3721	GAGAGTATCGAAACCGTTTCTTCTCATGATGATGATGATGATGATGATGATGATGATG	3780
DB	2641	TAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATGTTCCACTGGAAGCCAGA	2700	DB	3721	GAGAGTATCGAAACCGTTTCTTCTCATGATGATGATGATGATGATGATGATGATGATG	3780
QY	2701	TGACAAAGAAAGAAATACAGGAAAGAAAGAGAAAGAGTATGAGGAAATG	2760	QY	3781	ACAAATGAGGTTTGGACCCCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	3840
DB	2701	TGACAAAGAAAGAAATACAGGAAAGAAAGAGAAAGAGTATGAGGAAATG	2760	DB	3781	ACAAATGAGGTTTGGACCCCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	3840
QY	2761	CTGAAATACATAGGCGTTTCTGAAAGTGTGTTCTTGGAAAGAAAGCTCAGCATTTACAG	2820	QY	3841	CAATTTTCTGAGCAGTAAATGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
DB	2761	CTGAAATACATAGGCGTTTCTGAAAGTGTGTTCTTGGAAAGAAAGCTCAGCATTTACAG	2820	DB	3841	CAATTTTCTGAGCAGTAAATGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
QY	2821	ATAGTTACTTGGGCTTTTTCATGCGGAGAAAGAAAGTATTTCCAGATCTTCTGACG	2880	QY	3901	ACAAAGAAAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTA	3960
DB	2821	ATAGTTACTTGGGCTTTTTCATGCGGAGAAAGAAAGTATTTCCAGATCTTCTGACG	2880	DB	3901	ACAAAGAAAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTA	3960
QY	2881	AAGAGAGTCAATGAAGACAAATGGCATATCTTCACTGATAGCAAAATACCTGGAGGC	2940	QY	3961	CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGATGAT	4020
DB	2881	AAGAGAGTCAATGAAGACAAATGGCATATCTTCACTGATAGCAAAATACCTGGAGGC	2940	DB	3961	CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGATGAT	4020
				QY	4021	GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGCTTTTGTGGCATGAA	4080

[illegible]

D	b	2761	CTGAAATCA	CATAGGGCTT	ACTGAAGTTG	TCTTGGAGAAAG	CTGCGACAT	CACAG	2820
Q	y	2821	ATAGTTACT	TTGGGCTTTT	CCCATGGGAG	AAAAAAGTAT	TTCCAGATCT	TTCTCGACG	2880
D	b	2821	ATAGTTACT	TTGGGCTTTT	CCCATGGGAG	AAAAAAGTAT	TTCCAGATCT	TTCTCGACG	2880
Q	y	2881	AAGAGAGT	CATTGAAGAC	CAATTTGGCAT	PACTTCACTG	ATAGCAAAAT	TACTGGGAGGC	2940
D	b	2881	AAGAGAGT	CATTGAAGAC	CAATTTGGCAT	PACTTCACTG	ATAGCAAAAT	TACTGGGAGGC	2940
Q	y	2941	AACATAAAG	ATACATTTG	CAGATTCCT	CCAGATATG	TAAATAAAAT	TTCTAAATAGCAAGT	3000
D	b	2941	AACATAAAG	ATACATTTG	CAGATTCCT	CCAGATATG	TAAATAAAAT	TTCTAAATAGCAAGT	3000
Q	y	3001	TTGATTTCA	CATCGCGAAG	TGCCCTGCT	GCACATGCT	CACTGAT	TGACCGGATTTGTTA	3060
D	b	3001	TTGATTTCA	CATCGCGAAG	TGCCCTGCT	GCACATGCT	CACTGAT	TGACCGGATTTGTTA	3060
Q	y	3061	TGCAAGAAC	TGCAAGATAT	GTCTCCCTG	CAAGAAATTT	TGACAGAG	CGTCATTTTCACAAAAGTGC	3120
D	b	3061	TGCAAGAAC	TGCAAGATAT	GTCTCCCTG	CAAGAAATTT	TGACAGAG	CGTCATTTTCACAAAAGTGC	3120
Q	y	3121	GCCATTTCT	GAGGATATG	CAGTTTGC	CTTCTCTTAT	TTTATTTCT	CATGAGTGCAGTGC	3180
D	b	3121	GCCATTTCT	GAGGATATG	CAGTTTGC	CTTCTCTTAT	TTTATTTCT	CATGAGTGCAGTGC	3180
Q	y	3181	AGGCACAT	GAAATATAT	CTCAAGTCT	TTGATGAAG	TGTATACAGAT	CAATCTGCTGTTCTTGT	3240
D	b	3181	AGGCACAT	GAAATATAT	CTCAAGTCT	TTGATGAAG	TGTATACAGAT	CAATCTGCTGTTCTTGT	3240
Q	y	3241	CTGACAGAG	AAATCCCGA	AACTGGCTAC	CAAGAAATTC	ACGAAATTT	CGCGTTTAAAGTTTGCAGG	3300
D	b	3241	CTGACAGAG	AAATCCCGA	AACTGGCTAC	CAAGAAATTT	CGCGTTTAAAGTTTGCAGG	3300	
Q	y	3301	ATTTGACAG	GTCTGGAA	CACATGCTTA	TAAATTTGCT	CAAAAATGCT	TTCTGCTGATATCA	3360
D	b	3301	ATTTGACAG	GTCTGGAA	CACATGCTTA	TAAATTTGCT	CAAAAATGCT	TTCTGCTGATATCA	3360
Q	y	3361	CGCAGCTAA	ATAATATT	TCACCAACT	CAGAAATCCT	ACTATGAT	CCCAACCTGCCACCGG	3420
D	b	3361	CGCAGCTAA	ATAATATT	TCACCAACT	CAGAAATCCT	ACTATGAT	CCCAACCTGCCACCGG	3420
Q	y	3421	TCACTAAAG	CTAGTAA	CAAACTGTAA	ACCGATTA	CTGACAAAAT	TCGCAAGGATATA	3480
D	b	3421	TCACTAAAG	CTAGTAA	CAAACTGTAA	ACCGATTA	CTGACAAAAT	TCGCAAGGATATA	3480
Q	y	3481	AGGACAAA	ACAAATAT	ATAGTTTGA	ATCATCGG	AGAGAAATTC	GCCTTTTAAATATGA	3540
D	b	3481	AGGACAAA	ACAAATAT	ATAGTTTGA	ATCATCGG	AGAGAAATTC	GCCTTTTAAATATGA	3540
Q	y	3541	TTGCTACCA	ACCGTTTCT	CATGTGCT	GGTGGCAG	TGGATGAC	ATAGAAAAAACCCCTAGGA	3600
D	b	3541	TTGCTACCA	ACCGTTTCT	CATGTGCT	GGTGGCAG	TGGATGAC	ATAGAAAAAACCCCTAGGA	3600
Q	y	3601	AGTTTGTGT	TCCTGGA	TGACAAAT	GATGACCA	CAATCAT	TAAGATGCTCAGACAGTGAAGG	3660
D	b	3601	AGTTTGTGT	TCCTGGA	TGACAAAT	GATGACCA	CAATCAT	TAAGATGCTCAGACAGTGAAGG	3660
Q	y	3661	CTGTCTCAG	GGAATCT	TATGAATCC	ATGTTCC	CCCATACCT	TTCCCAATTTGAACTGC	3720
D	b	3661	CTGTCTCAG	GGAATCT	TATGAATCC	ATGTTCC	CCCATACCT	TTCCCAATTTGAACTGC	3720
Q	y	3721	GAGAGTAT	CGAAACCG	TTTCCCT	CATATG	CATGAGCT	GCAGGAATGAGGGCTTTATCGAG	3780
D	b	3721	GAGAGTAT	CGAAACCG	TTTCCCT	CATATG	CATGAGCT	GCAGGAATGAGGGCTTTATCGAG	3780
Q	y	3781	ACAAATGAA	GTGTTTGG	ACCCATTTG	TACTAG	CAACAT	GATATGTTTACTATATTTCT	3840
D	b	3781	ACAAATGAA	GTGTTTGG	ACCCATTTG	TACTAG	CAACAT	GATATGTTTACTATATTTCT	3840
Q	y	3841	CATTTTGTG	TGAGAGT	TTAAATGCA	CTTAAG	CGGAAGATAT	TTTCCGACAGAGAGATAC	3900
D	b	3841	CATTTTGTG	TGAGAGT	TTAAATGCA	CTTAAG	CGGAAGATAT	TTTCCGACAGAGAGATAC	3900

XX	US2003124653-A1.
FN	XX
XX	XX
PD	03-JUL-2003.
XX	XX
PF	21-DEC-2001; 2001US-00023890.
XX	XX
PR	21-DEC-2001; 2001US-00023890.
XX	XX
PA	(NOVA-) NOVAZYNE PHARM INC.
XX	XX
PI	Canfield WM;
XX	XX
DR	WPI; 2003-810985/76.
DR	P-PSDB; ABW01537, ABW01538.
XX	XX
PT	Producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.
PT	Claim 12; Page 13-15; 46pp; English.
XX	XX
CC	The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha/beta subunit precursor cDNA
XX	XX
SQ	Sequence 5597 BP; 1678 A; 1151 C; 1219 G; 1549 T; 0 U; 0 Other;
	Query Match 100.0%; Score 5597; DB 10; Length 5597;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 CGAGCGCAGCGGGCGTCGTGCOCGAGCTGCATAGCGCGGCCCGGAGGCTGTGACC 60
Db	1 CGAGCGCAGCGGGCGTCGTGCOCGAGCTGCATAGCGCGGCCCGGAGGCTGTGACC 60
Qy	61 TCGCGCGCGCGCCGACGCGGGCCCCTGAATGCGCGCTCGCTGAGCGCGCGCGCGCGC 120
Db	61 TCGCGCGCGCGCCGACGCGGGCCCCTGAATGCGCGCTCGCTGAGCGCGCGCGCGCGC 120
Qy	121 GC CGCGCTCAGGTCCTCGCGGCGTGCGGTGCCACAGCATGGCGTCTCAAGCTCC 180
Db	121 GC CGCGCTCAGGTCCTCGCGGCGTGCGGTGCCACAGCATGGCGTCTCAAGCTCC 180
Qy	181 TGCAGAGACAACCTATACCTGCCTGCCACAGCATGGCGTCTCAAGCTCTCTTGG 240
Db	181 TG CAGAGACAACCTATACCTGCCTGCCACAGCATGGCGTCTCAAGCTCTCTTGG 240
Qy	241 GCCTCGTTGTCAACCATCGTCTCCGCGTTCAGATTCCGAGAGTGGTTCGGAATGAGCC 300
Db	241 GCCTCGTTGTCAACCATCGTCTCCGCGTTCAGATTCCGAGAGTGGTTCGGAATGAGCC 300
Qy	301 GAGATCAATACGANTTTTTTGTGATTCCTATAGACAATATTCTGTGAAAAGTCCTTTTC 360
Db	301 GAGATCAATACCATGTTTTTTGTGATTCCTATAGACAATATTCTGTGAAAAGTCCTTTTC 360
Qy	361 AGAATCGCTTTGTCTGCCATGCCGATTGACGTTGTTTACACTGGGTCAATGGCACAG 420
Db	361 AGAATCGGCTTTGTCTGCCATTECCGATTGACGTTGTTTACACTGGGTGAAATGGCACAG 420
Qy	421 ATCTTGAATCTATGAGGAATCTACGACGGTCAGAGAACAGATGGAGGAGGACAGAAAG 480
Db	421 ATCTTGAATCTATGAGGAATCTACGACGGTCAGAGAACAGATGGAGGAGGACAGAAAG 480
Qy	481 CAATGAGANAATCTTGGGAAAAACAACGGAACCTACTTAAGAGAGGTGAGAAGCACT 540
Db	481 CAATGAGAGAAATCTTGGGAAAAACAACGGAACCTACTTAAGAGAGGTGAGAAGCACT 540
Qy	541 TAGAGTGTGTGTTAAACACACTGCATTAAGGTGCCAATGCTTTGACTAGCAACCCCTTCG 600

Db 1621 GTATTGGAGTTGGACACCCCTTGGCGAGTTTGGTGGAGGAAATAAACAAGTGTCTCTTACTGTA 1680
Qy 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740
Db 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740
Qy 1741 TGTCTGTGGTGTGATCTGGCGACTGTGGCGAGATCAATTTTCATGAATGTGATAAG 1800
Db 1741 TGTCTGTGGTGTGATCTGGCGACTGTGGCGAGATCAATTTTCATGAATGTGATAAG 1800
Qy 1801 TGATCTTCTCCCAACAGACTCACTATATTTTCCAAAGAGTGAATGCTGCTTATT 1860
Db 1801 TGATCTTCTCCCAACAGACTCACTATATTTTCCAAAGAGTGAATGCTGCTTATT 1860
Qy 1861 TCAGCTTTGCGAAGTAGCCAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCAATA 1920
Db 1861 TCAGCTTTGCGAAGTAGCCAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCAATA 1920
Qy 1921 TTCCGACATGCTTCTATTTGCCAACCAAGTGGAAACCATCCACCTCATATGACACAGTGGAA 1980
Db 1921 TTCCGACATGCTTCTATTTGCCAACCAAGTGGAAACCATCCACCTCATATGACACAGTGGAA 1980
Qy 1981 TGAATGCCACCAATACATTTTAACTTCAAGTTTCAAAATACAAACGATGAAGATTCA 2040
Db 1981 TGAATGCCACCAATACATTTTAACTTCAAGTTTCAAAATACAAACGATGAAGATTCA 2040
Qy 2041 AAATGSCAGATAACAGTGGAGTGGACCAAGGAGGGACCAAACTCAATTTCTACGGCCC 2100
Db 2041 AAATGSCAGATAACAGTGGAGTGGACCAAGGAGGGACCAAACTCAATTTCTACGGCCC 2100
Qy 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCAATACACTTTCTCCAGAGGCGGAAATCCTTT 2160
Db 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCAATACACTTTCTCCAGAGGCGGAAATCCTTT 2160
Qy 2161 TTGAGGATATCCCAAGAAAACGCTTCCCGAGTTTGAAGNATGATGTTAACTCAA 2220
Db 2161 TTGAGGATATCCCAAGAAAACGCTTCCCGAGTTTGAAGNATGATGTTAACTCAA 2220
Qy 2221 CAAGGAGAGCCAGGAAGAGTGAATAATCCCTCGTGAATAATTTCTACTCTTCCAAAG 2280
Db 2221 CAAGGAGAGCCAGGAAGAGTGAATAATCCCTCGTGAATAATTTCTACTCTTCCAAAG 2280
Qy 2281 ACGCCAGTCTCAATACCTTGGATTTGCAACTGGAACTGAGACATCACTTTGA 2340
Db 2281 ACGCCAGTCTCAATACCTTGGATTTGCAACTGGAACTGAGACATCACTTTGA 2340
Qy 2341 AAGGATACAAATTTGTCGAAGTCAAGCTTGTGAGATCAATTTCTGATGAATCTCAGCATG 2400
Db 2341 AAGGATACAAATTTGTCGAAGTCAAGCTTGTGAGATCAATTTCTGATGAATCTCAGCATG 2400
Qy 2401 CTAAATTAATAATCAAGCTTAATAACAGATGAAAACAAATGACAGTTTGGTGGCTCCAC 2460
Db 2401 CTAAATTAATAATCAAGCTTAATAACAGATGAAAACAAATGACAGTTTGGTGGCTCCAC 2460
Qy 2461 AGGAAAACAGGTTCAATAAAGACATTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
Db 2461 AGGAAAACAGGTTCAATAAAGACATTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
Qy 2521 AGAGTTTGAATTTCTCGATGAGTGTAAAGTGAATGTTGATGACAGGGTCAAGATC 2580
Db 2521 AGAGTTTGAATTTCTCGATGAGTGTAAAGTGAATGTTGATGACAGGGTCAAGATC 2580
Qy 2581 CACCCCTGGATTCGAGACCAAGAGATTTAGAGTGGAACTCACACCCAAACCA 2640
Db 2581 CACCCCTGGATTCGAGACCAAGAGATTTAGAGTGGAACTCACACCCAAACCA 2640
Qy 2641 TAGCGGAAATGTGACAAAAGAAAAGCCGCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700
Db 2641 TAGCGGAAATGTGACAAAAGAAAAGCCGCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700
Qy 2701 TGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAGACAGTAGATGGAGGAAAATG 2760
Db 2701 TGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAGACAGTAGATGGAGGAAAATG 2760

Qy 2761 CTGAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAGAAAAGCTGCGAGATTACACAG 2820
Db 2761 CTGAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAGAAAAGCTGCGAGATTACACAG 2820
Qy 2821 ATAGTTACTTTGGCTTTTGGCCATGGGAGAAAAGTATTTCCAGATCTTCTCGACG 2880
Db 2821 ATAGTTACTTTGGCTTTTGGCCATGGGAGAAAAGTATTTCCAGATCTTCTCGACG 2880
Qy 2881 AAGAAGAGTCATGTAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC 2940
Db 2881 AAGAAGAGTCATGTAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC 2940
Qy 2941 AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATAAATTTCTAATAGCAAGT 3000
Db 2941 AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATAAATTTCTAATAGCAAGT 3000
Qy 3001 TTGGATTTCATCGCGGAAAGTCCCTGCTCAGATCGCTCAGATGTTGACCGGATTGTTA 3060
Db 3001 TTGGATTTCATCGCGGAAAGTCCCTGCTCAGATCGCTCAGATGTTGACCGGATTGTTA 3060
Qy 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTCAAAAGTGC 3120
Db 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTCAAAAGTGC 3120
Qy 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTATTTTATTTATCTCATGATGTCAGTGC 3180
Db 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTATTTTATTTATCTCATGATGTCAGTGC 3180
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240
Qy 3241 CTGACAGAGAAATCGAACAACCTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG 3300
Db 3241 CTGACAGAGAAATCGAACAACCTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG 3300
Qy 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTCAAAAATGCTTCTCTCTGATATCA 3360
Db 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTCAAAAATGCTTCTCTCTGATATCA 3360
Qy 3361 CGCAGCTAAAATAATTTCCCAACTCAGGAATCTCTATGATCCCAACTGCCACCGG 3420
Db 3361 CGCAGCTAAAATAATTTCCCAACTCAGGAATCTCTATGATCCCAACTGCCACCGG 3420
Qy 3421 TCACCTAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATCCACAAGCATATA 3480
Db 3421 TCACCTAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATCCACAAGCATATA 3480
Qy 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGA 3540
Db 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAGAGAGAAATCGCTTTTAAATGA 3540
Qy 3541 TTCGTACCAACGTTTCTCATGTTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGA 3600
Db 3541 TTCGTACCAACGTTTCTCATGTTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGA 3600
Qy 3601 AGTTTGTGTGCTGAATGACAAATTTGACCAATCATATAAGATGCTCAGACAGTGAAGG 3660
Db 3601 AGTTTGTGTGCTGAATGACAAATTTGACCAATCATATAAGATGCTCAGACAGTGAAGG 3660
Qy 3661 CTGTTCTCAGGAGCTTCTATGAATCCCATACCTTCCCAATTTTGAATGCTGCAAT 3720
Db 3661 CTGTTCTCAGGAGCTTCTATGAATCCCATGTTTCCCATACCTTCCCAATTTTGAATGCTGCAAT 3720
Qy 3721 CAGAGTATCGAAACCGTTTCTCTCATATGATGCTGAGGAAATGGAGGCTTATCGAG 3780
Db 3721 CAGAGTATCGAAACCGTTTCTCTCATATGATGCTGAGGAAATGGAGGCTTATCGAG 3780
Qy 3781 ACAAATTGAAGTTTGGACCCATTTGCTAGTACCAATGATTGATTATCTTCTATATCT 3840
Db 3781 ACAAATTGAAGTTTGGACCCATTTGCTAGTACCAATGATTGATTATCTTCTATATCT 3840

QY	1532	TAATAAATCAGCCTCGGATTTGGGATGTTGGGATTTGCTCTGGAAACAGTGGAGGAGTCG	1591
DB	1501	TAATAAATCAGCCTCGGATTTGGGATGTTGGGATTTGCTCTGGAAACAGTGGAGGAGTCG	1560
QY	1592	CTATATTCAGAGGTGAGGTGACCTGGAGTATTCGAGTGTGACACACCCCTCGCATTTGG	1651
DB	1561	CTATATTCAGAGGTGAGGTGACCTGGAGTATTCGAGTGTGACACACCCCTCGCATTTGG	1620
QY	1652	TGAGGAAATAACAGTGTCTCTTACTGTAAATCAGGGATGTGCGAATTCCTGCTCGCTGA	1711
DB	1621	TGAGGAAATAACAGTGTCTCTTACTGTAAATCAGGGATGTGCGAATTCCTGCTCGCTGA	1680
QY	1712	TAAAGTCTGTGACCAAGCATGCAATGTCTTGTCTGTGGTGTGATGTGCGCATGTGG	1771
DB	1681	TAAAGTCTGTGACCAAGCATGCAATGTCTTGTCTGTGGTGTGATGTGCGCATGTGG	1740
QY	1772	GCAAGATCATTTTTCATGATGTTAAAGTGTATCTTCTCCCAACAGACATCACTATAT	1831
DB	1741	GCAAGATCATTTTTCATGATGTTAAAGTGTATCTTCTCCCAACAGACATCACTATAT	1800
QY	1832	TATTCCTCAAGGTGAATGCTGCTTATTTTTCAGCTTTTGCAGAGTAGCCAAAGAGGAGT	1891
DB	1801	TATTCCTCAAGGTGAATGCTGCTTATTTTTCAGCTTTTGCAGAGTAGCCAAAGAGGAGT	1860
QY	1892	TGAAGGTGCTATAGTGACATCCAAATTAATTGACATGCTTCTTATTCGCAACAGAGTGAA	1951
DB	1861	TGAAGGTGCTATAGTGACATCCAAATTAATTGACATGCTTCTTATTCGCAACAGAGTGAA	1920
QY	1952	AACCATCAACCTCAATAATGACAGTGAATGAATGCCACCAACAATTAATTTTCTCAC	2011
DB	1921	AACCATCAACCTCAATAATGACAGTGAATGAATGCCACCAACAATTAATTTTCTCAC	1980
QY	2012	GTTCCTCAAAATCAAAACGATGAAGTTCAAAATGCAGATTAACAGTGGAGGTGGACAAAG	2071
DB	1981	GTTCCTCAAAATCAAAACGATGAAGTTCAAAATGCAGATTAACAGTGGAGGTGGACAAAG	2040
QY	2072	GGAGGACCAAACTGAAATCTTACGCGCCAGAGGTTACGAAATTTAGTTAGTCCCAT	2131
DB	2041	GGAGGACCAAACTGAAATCTTACGCGCCAGAGGTTACGAAATTTAGTTAGTCCCAT	2100
QY	2132	AACACTTTCTCAGAGGCGGAAATCCTTTTGGAGATATTCCTCAAGAAACAGCTTCCC	2191
DB	2101	AACACTTTCTCAGAGGCGGAAATCCTTTTGGAGATATTCCTCAAGAAACAGCTTCCC	2160
QY	2192	GAAATTTAAGACATGATGTTAACTCAACAGGAGGCCAGGAGGTGAAATTTCC	2251
DB	2161	GAAATTTAAGACATGATGTTAACTCAACAGGAGGCCAGGAGGTGAAATTTCC	2220
QY	2252	CCTGGTAAATATTTCACTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTGGATTT	2311
DB	2221	CCTGGTAAATATTTCACTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTGGATTT	2280
QY	2312	GCACTGGAAACATGAGACATCCTTTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGCT	2371
DB	2281	GCACTGGAAACATGAGACATCCTTTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGCT	2340
QY	2372	GAGATCATTTCTGATGAACTCACAGCATGCTAAATTAATAATCAAGCTATAATAACAGA	2431
DB	2341	GAGATCATTTCTGATGAACTCACAGCATGCTAAATTAATAATCAAGCTATAATAACAGA	2400
QY	2432	TGAAACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTCATAAAGCATCTTGCC	2491
DB	2401	TGAAACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTCATAAAGCATCTTGCC	2460
QY	2492	AAACAGCTTAGAGTGTCTGAAAGATTCAGAGGTTGACTTTTCTGCGAGTGAATAA	2551
DB	2461	AAACAGCTTAGAGTGTCTGAAAGATTCAGAGGTTGACTTTTCTGCGAGTGAATAA	2520
QY	2552	AGTGAATGCTATGACCAAGGTTCAGAAATCCACCCCTGGACTTGGAGACCAAGAGATT	2611
DB	2521	AGTGAATGCTATGACCAAGGTTCAGAAATCCACCCCTGGACTTGGAGACCAAGAGATT	2580
QY	2612	TAGAGTGGAACTCACACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAGCCCC	2671
DB	2581	TAGAGTGGAACTCACACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAGCCCC	2640
QY	2672	ATCTCTGATTTGTTCCATCTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGA	2731
DB	2641	ATCTCTGATTTGTTCCATCTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGA	2700
QY	2732	AAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGGGTTTACTGAACTGTT	2791
DB	2701	AAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGGGTTTACTGAACTGTT	2760
QY	2792	ACTTGGAGAAAGCTGCGAGCATTTACAGATAGTTACTTTGGGCTTTTGGCCATGGAGAA	2851
DB	2761	ACTTGGAGAAAGCTGCGAGCATTTACAGATAGTTACTTTGGGCTTTTGGCCATGGAGAA	2820
QY	2852	AAAAAGTATTTTCCAAAGATCTTCTCGACGAAGAGAGTCAATTCGAACACAAATTCGCATA	2911
DB	2821	AAAAAGTATTTTCCAAAGATCTTCTCGACGAAGAGAGTCAATTCGAACACAAATTCGCATA	2880
QY	2912	CTTCACTGATPAGCAAAATACTGCGGAGGCAACTTAAAGATACATTTTGCAGATTTCCCTCAG	2971
DB	2881	CTTCACTGATPAGCAAAATACTGCGGAGGCAACTTAAAGATACATTTTGCAGATTTCCCTCAG	2940
QY	2972	ATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCA	3031
DB	2941	ATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCA	3000
QY	3032	CATGCTCTCATGATTTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTTCCCTGAAGA	3091
DB	3001	CATGCTCTCATGATTTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTTCCCTGAAGA	3060
QY	3092	ATTGTGACAGACGTCAATTTTCAAAAGTGGCCATTTCTGAGGATATGCGATTTGCTTCTC	3151
DB	3061	ATTGTGACAGACGTCAATTTTCAAAAGTGGCCATTTCTGAGGATATGCGATTTGCTTCTC	3120
QY	3152	TTATTTTATTTATCTCATGATGCTGACGACCTGCAATATATCTCAAGTCTTTTGCATGA	3211
DB	3121	TTATTTTATTTATCTCATGATGCTGACGACCTGCAATATATCTCAAGTCTTTTGCATGA	3180
QY	3212	AGTTGATACAGATCAATCTGTTGCTTTGCTGACAGAGAAATCCGAAACCTGCTACCAG	3271
DB	3181	AGTTGATACAGATCAATCTGTTGCTTTGCTGACAGAGAAATCCGAAACCTGCTACCAG	3240
QY	3272	AATTCAGAACCTGCTTAAAGTTTGCAGGATTTGACAGCTCTGGRACACATGCTATATAA	3331
DB	3241	AATTCAGAACCTGCTTAAAGTTTGCAGGATTTGACAGCTCTGGRACACATGCTATATAA	3300
QY	3332	TTGCTCAAAAATGCTTCTGCTGATATCAGCAGCTTAAATAATATTTCCACCACTCAGGA	3391
DB	3301	TTGCTCAAAAATGCTTCTGCTGATATCAGCAGCTTAAATAATATTTCCACCACTCAGGA	3360
QY	3392	ATCCTACTATGATCCCAACCTGCGCCGCTCACTAAAGTCTAGTAAACAACTGTAAACC	3451
DB	3361	ATCCTACTATGATCCCAACCTGCGCCGCTCACTAAAGTCTAGTAAACAACTGTAAACC	3420
QY	3452	AGTAACTGACAAATCCACAAAGCATATAAGGACAAAAACAAATATAGGTTTGAATTCAT	3511
DB	3421	AGTAACTGACAAATCCACAAAGCATATAAGGACAAAAACAAATATAGGTTTGAATTCAT	3480
QY	3512	GGGAGAGAGAAATCGCTTTTAAATGATTTGTCACCAAGTTTCTCATGTGTTGGCCCA	3571
DB	3481	GGGAGAGAGAAATCGCTTTTAAATGATTTGTCACCAAGTTTCTCATGTGTTGGCCCA	3540
QY	3572	GTTGGATGACATAAGAAAAACCTAGGAAGTTTGTTCCTGGAATGACAACTTGCACCA	3631
DB	3541	GTTGGATGACATAAGAAAAACCTAGGAAGTTTGTTCCTGGAATGACAACTTGCACCA	3600
QY	3632	CAATCATAAAGATGCTCAGACAGTGAAGCTGTTTCTCAGGACCTTCTATGAATCCATGTT	3691
DB	3601	CAATCATAAAGATGCTCAGACAGTGAAGCTGTTTCTCAGGACCTTCTATGAATCCATGTT	3660
QY	3692	CCCCATACCTTCCCAATTTGAATGCGCAAGAGAGTATCGAAACCGTTTCTTCCATATGCA	3751

FT	mat_peptide	243..3728
FT	FT	/*tag= c
FT	FT	/product= "Mature human NAAP-13"
XX	XX	
XX	W0200250279-A2.	
XX	XX	
XX	27-JUN-2002.	
XX	XX	
XX	19-DEC-2001; 2001WO-US050256.	
XX	XX	
PR	21-DEC-2000; 2000US-0257714P.	
PR	05-JAN-2001; 2001US-0260081P.	
PR	16-JAN-2001; 2001US-0262302P.	
PR	23-JAN-2001; 2001US-0263823P.	
PR	02-FEB-2001; 2001US-0266089P.	
PR	29-OCT-2001; 2001US-0348442P.	
XX	XX	
XX	(INCY-) INCVTE GENOMICS INC.	
XX	XX	
XX	Baughin MR, Lu Y, Arvizu C, Rankumar J, Yao MG, Policky JL;	
PI	Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;	
PI	Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;	
PI	Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalla AJA;	
PI	Lee EA, Yang J, Gorvad AB, Emerling BM, Marquis JP, Lee SY;	
PI	Swarnakar A, Reddy R;	
XX	XX	
XX	XX	
DR	WPI; 2002-519887/55.	
DR	P-PSDB; AA255294.	
XX	XX	
PT	Nucleic acid associated proteins and nucleic acids for diagnosing,	
PT	treating and preventing cell proliferative (e.g. cancers), neurological	
PT	(e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).	
XX	XX	
PS	Claim 84; Page 189-190; 193pp; English.	
XX	XX	
CC	The invention relates to nucleic acid-associated proteins (NAAP) and	
CC	nucleic acids. The nucleic acid and amino acid sequences are useful for	
CC	diagnosing, treating and preventing cell proliferative e.g.	
CC	arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological	
CC	(e.g. epilepsy, Alzheimer's disease or stroke), developmental, and	
CC	autoimmune disorders (e.g. AIDS, allergies or anaemia) or infections	
CC	(e.g. malaria, or leishmania), as well as in assessing the effects of	
CC	exogenous compound on the expression of nucleic acid and amino acid	
CC	sequences of nucleic acid-associated proteins. The invention is useful in	
CC	gene therapy. The present sequence is human NAAP-13 CDNA	
XX	XX	
XX	Sequence 4277 BP; 1125 A; 891 C; 925 G; 1135 T; 0 U; 1 Other;	

Query Match	72.5%;	Score 4057;	DB 6;	Length 4277;
Best Local Similarity	95.7%;	Pred. No. 0;		
Matches 4264;	Conservative	0;	Mismatches 11;	Indels 181; Gaps 2
Qy	26	GGAGCTGCAATGAGCGCGCCGCGAGGCTGTGACCTCGCGCGCGCGCGCCGACCGCGGGCC	85	
Db	2	GGAGCTGCAATGAGCGCGCCGCGAGGCTGTGACCTCGCGCGCGCGCGCCGACCGGGCC	61	
Qy	86	CTTGAAATGGCGGCTCGCTGAGCGCGCGCGCGCGCGCGCTCAGGCTCCTCGGGGCGTG	145	
Db	62	CTTGAAATGGCGGCTCGCTGAGCGCGCGCGCGCGCGCTCAGGCTCCTCGGGGCGTG	121	
Qy	146	GGGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCCTCGAGAGACAACCTATACCTGCCT	205	
Db	122	GGGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCCTCGAGAGACAGACCTATACCTGCCT	181	
Qy	206	GTCCACACAGTATGGGCTCTACGTGTGCTTCTTGGGCGTGGTTGTACACATCGTCTCCGC	265	
Db	182	GTCCACACAGTATGGGCTCTACGTGTGCTTCTTGGGCGTGGTTGTACACATCGTCTCCGC	241	
Qy	266	CTTCAGATTTCGAGAGAGTGGTTCTGGGAATGAGCGCGAGATCAATACCATGTTTGTTTGA	325	
Db	242	CTTCAGATTTCGAGAGAGTGGTTCTGGGAATGAGCGCGAGATCAATACCATGTTTGTTTGA	301	
Qy	326	TTCTCTATGAGACAAATATGCTGCGAAAGTCCTTTTCAAGATCGGCTTTGTGTGCCCAATGCC	385	

QY 2546 TGTAAGAGTGAATGGTTCATGACACAGGCTCAGAAATCCACCCCTGGACTTTGGAGACCACAGC 2605
Db 2342 TGTAAGAGTGAATGGTTCATGACACAGGCTCAGAAATCCACCCCTGGACTTTGGAGACCACAGC 2401
QY 2606 AAGATTTAGAGTGAAACTCACACCCAAAATAAGGCGGAAATGTGACAAAAGAAA 2665
Db 2402 AAGATTTAGAGTGAAACTCACACCCAAAATAAGGCGGAAATGTGACAAAAGAAA 2461
QY 2666 GCCCCCATCTCTGATTTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAATACAGG 2725
Db 2462 GCCCCCATCTCTGATTTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAATACAGG 2521
QY 2726 GAAAGAAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGCGCTTACTGA 2785
Db 2522 GAAAGAAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGCGCTTACTGA 2581
QY 2786 AGTGTACTTGGAGAGAAAGCTGAGCAATACACAGATAGTTACTTGGGCTTTTGGCATG 2845
Db 2582 AGTGTACTTGGAGAGAAAGCTGAGCAATACACAGATAGTTACTTGGGCTTTTGGCATG 2641
QY 2846 GGAGAAAAGAAAGTATTTCCAGATCTTCTCGAGGAAAGAGAGTCAATTGAAGACACAATT 2905
Db 2642 GGAGAAAAGAAAGTATTTCCAGATCTTCTCGAGGAAAGAGAGTCAATTGAAGACACAATT 2701
QY 2906 GGCATACCTCTGATAGCAAAAATATCTGGAGGCAACTAAAAGATACATTTGCAGATTC 2965
Db 2702 GGCATACCTCTGATAGCAAAAATATCTGGAGGCAACTAAAAGATACATTTGCAGATTC 2761
QY 2966 CCTCAGATATGAATTAATTAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCC 3025
Db 2762 CCTCAGATATGAATTAATTAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCC 2821
QY 3026 TGCTCACATGCTCACATGATGTACCGGATTTGATGCAAGAACTGCAAGATATGTTCCC 3085
Db 2822 TGCTCACATGCTCACATGATGTACCGGATTTGATGCAAGAACTGCAAGATATGTTCCC 2881
QY 3086 TGAAGAAATTCACAGAGCTCAATTCACAAAGTGGCCATCTCAGGATATGCAAGTTGC 3145
Db 2882 TGAAGAAATTCACAGAGCTCAATTCACAAAGTGGCCATCTCAGGATATGCAAGTTGC 2941
QY 3146 CTTCCTCTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTT 3205
Db 2942 CTTCCTCTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTT 3001
QY 3206 TGATGAAGTGAATACAGATCAATCTGGTGCTTGTCTGACAGAGAAATCCGAACATGGC 3265
Db 3002 TGATGAAGTGAATACAGATCAATCTGGTGCTTGTCTGACAGAGAAATCCGAACATGGC 3061
QY 3266 TACAGAAATTCACAGAACTGCGGTTAAGTTTGCAGATTTGCAGAGTCTGGAACACATGCT 3325
Db 3062 TACAGAAATTCACAGAACTGCGGTTAAGTTTGCAGATTTGCAGAGTCTGGAACACATGCT 3121
QY 3326 AATAAATGCTCAAAAATGCTTCTGCTGATPATCACGAGCTAAATAATATTCACCAAC 3385
Db 3122 AATAAATGCTCAAAAATGCTTCTGCTGATPATCACGAGCTAAATAATATTCACCAAC 3181
QY 3386 TCAGGAATCTCTATGATCCCACTGCGCACTGCGCTCACTAAAGTCTAGTAACAACCTG 3445
Db 3182 TCAGGAATCTCTATGATCCCACTGCGCACTGCGCTCACTAAAGTCTAGTAACAACCTG 3241
QY 3446 TAAACAGTAAGTACAAAATCCAAAGCATATAAGGACAAAACAAAATATAGTTTGA 3505
Db 3242 TAAACAGTAAGTACAAAATCCAAAGCATATAAGGACAAAACAAAATATAGTTTGA 3301
QY 3506 AATCATGGAGAGAGAAATCGCTTTTAAATGATTCGACCAACGTTTCTCATGTGGT 3565
Db 3302 AATCATGGAGAGAGAAATCGCTTTTAAATGATTCGACCAACGTTTCTCATGTGGT 3361
QY 3566 TGGCCAGTTGATGACATAAGAAAACCCCTAGGAAGTTGTTTGGCTGTAATGACACAT 3625
Db 3362 TGGCCAGTTGATGACATAAGAAAACCCCTAGGAAGTTGTTTGGCTGTAATGACACAT 3421

QY 3626 TGACCACAATCATAAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGACCTTCTATGAATC 3685
Db 3422 TGACCACAATCATAAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGACCTTCTATGAATC 3481
QY 3686 CATGTTCCCACTACCTTCCCAATTTGAACTGCCAAGAGAGATCGAAACCGTTTCCCTTCA 3745
Db 3482 CATGTTCCCACTACCTTCCCAATTTGAACTGCCAAGAGAGATCGAAACCGTTTCCCTTCA 3541
QY 3746 TATGATCAGCTGCGAGGAATGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATG 3805
Db 3542 TATGATCAGCTGCGAGGAATGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATG 3601
QY 3806 TGTACTAGCAACATGATATGTTTACTATATCTCATTTCTGCTGAGCAGTTAAATGTC 3865
Db 3602 TGTACTAGCAACATGATATGTTTACTATATCTCATTTCTGCTGAGCAGTTAAATGTC 3661
QY 3866 ACTTAAGCGGAAGATATTTCCAGAGAGGAGATACAAAGAAAGCTAGTCCCAATCGAAT 3925
Db 3662 ACTTAAGCGGAAGATATTTCCAGAGAGGAGATACAAAGAAAGCTAGTCCCAATCGAAT 3721
QY 3926 CAGATATAGAAAGATCTTCAATTTGAAAACCATCTACCTCAGCATTTACTGAGCAATTTAA 3985
Db 3722 CAGATATAGAAAGATCTTCAATTTGAAAACCATCTACCTCAGCATTTACTGAGCAATTTAA 3781
QY 3986 RACTCAGCTTCACAGAGATGCTTTGCTGATGCTTACAGTTTGGCCCGGAAAGG 4045
Db 3782 RACTCAGCTTCACAGAGATGCTTTGCTGATGCTTACAGTTTGGCCCGGAAAGG 3841
QY 4046 AAAATATCAGTACCATGCTGTTTGTGGCAATATAGCCCACTGACTAGCAATTTAT 4105
Db 3842 AAAATATCAGTACCATGCTGTTTGTGGCAATATAGCCCACTGACTAGCAATTTAT 3901
QY 4106 TAACCAACCCACTGAAAATTTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGA 4165
Db 3902 TAACCAACCCACTGAAAATTTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGA 3961
QY 4166 ATTTGCTCATGACCTGTCATCCCTTTTATAAAGGCTCACTGACAGAGACAGCTGTT 4225
Db 3962 ATTTGCTCATGACCTGTCATCCCTTTTATAAAGGCTCACTGACAGAGACAGCTGTT 4021
QY 4226 AATTTCCACAGCAATCATTCGAGCACTAATTTAGGAGAGCCTATGCGAGCTGGG 4285
Db 4022 AATTTCCACAGCAATCATTCGAGCACTAATTTAGGAGAGCCTATGCGAGCTGGG 4081
QY 4286 GTGATTTGAGAGGCTCCAGTCTTTGCAATTCGAAGGCTTTTGTAAAGTTTGCAT 4345
Db 4082 GTGATTTGAGAGGCTCCAGTCTTTGCAATTCGAAGGCTTTTGTAAAGTTTGCAT 4141
QY 4346 TTTTCTTTTCAATTTCCCAATTTTAAAGTATTAAGTAACTAGTTATTTCTGCTGCTG 4405
Db 4142 TTTTCTTTTCAATTTCCCAATTTTAAAGTATTAAGTAACTAGTTATTTCTGCTGCTG 4201
QY 4406 AGTATAAGAAATGGAGTCTTAAA-CCTATTTTATAGATGTTATTTTAAATTAATGACG 4464
Db 4202 AGTATAAGAAATGGAGTCTTAAACCTATTTTATAGATGTTATTTTAAATTAATGACG 4261
QY 4465 AATATCACCTCTTATT 4480
Db 4262 ATTTNACCTCTTTT 4277

RESULT 7

ACC81010
ID ACC81010 standard; cDNA; 3783 BP.

XX ACC81010;

XX AC

DT 01-AUG-2003 (first entry)

XX Human GlcNAc-phosphotransferase related cDNA SEQ ID NO: 20.

XX Human; N-acetylglucosamine-1-phosphotransferase; nephroretic;

KW GlcNAc-phosphotranferase; phosphodiester alpha-GlcNAcase;

N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase; enzyme replacement therapy; phosphorylated lysosomal hydrolase; lysosomal storage disease; gene; ss.

Homo sapiens.

US6537785-B1.

XX 25-MAR-2003.

XX 10-AUG-2000; 2000US-00636077.

XX 14-SEP-1999; 99US-0153831P.

PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX Canfield WN;

XX WPI; 2001-290356/30.

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

Example 9; Page 52-54; 62pp; English.

The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-1-phosphotransferase) (I) and phosphodiester alpha-galactosidase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a patient suffering from a lysosomal storage disease. The present sequence is used in the exemplification of the invention

XX Sequence 3783 BP; 1198 A; 776 C; 835 G; 974 T; 0 U; 0 Other;

Query Match 57.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Fred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 165 ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCTCCACAGGTATGGGCTC 224
DB 7 ATGGGTTCAAGCTCTTCAGAGACAAACCTATACCTGCTCCACAGGTATGGGCTC 66

QY 225 TACGTGCTCTTGGGGCTGTGTGTCACATCGTCTCCGCTTCCAGTTCGGAGAGGTG 284
DB 67 TACGTGCTCTTGGGGCTGTGTGTCACATCGTCTCCGCTTCCAGTTCGGAGAGGTG 126

QY 285 GTTCTGGAATGAGCGGAGATCAATACCATGTTTGTGTTGATTCTCTATAGACAAATTT 344
DB 127 GTTCTGGAATGAGCGGAGATCAATACCATGTTTGTGTTGATTCTCTATAGACAAATTT 186

QY 345 GCTGAAAGTCTTTTCAAGATCGGCTTGTCTGCCCATGCGATGACGTTGTTTACACC 404
DB 187 GCTGAAAGTCTTTTCAAGATCGGCTTGTCTGCCCATGCGATGACGTTGTTTACACC 246

QY 405 TGGGTGATGACAGATCTTGAATCTGAACTGAAAGCACTACAGCAGGTTCAGAGACAGATG 464
DB 247 TGGGTGATGACAGATCTTGAATCTGAACTGAAAGCACTACAGCAGGTTCAGAGACAGATG 306

QY 465 GAGGAGGAGCAAGAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTTACTAAG 524
DB 307 GAGGAGGAGCAAGAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTTACTAAG 366

QY 525 AAGAGTGAGAGCAGTTAGAGTGTGCTTAACACACTGCACTAAGTGGCCATGCTTGTA 584
DB 367 AAGAGTGAGAGCAGTTAGAGTGTGCTTAACACACTGCACTAAGTGGCCATGCTTGTA 426

QY 585 CTGGACCCAGCCCTGCCAGCAACATCACCCCTGAAGGACGTGCCATCTCTTATCTCTCT 644

DB 427 CTGGACCCAGCCCTGCCAGCAACATCACCCCTGAAGGACCTGCATCTCTTTATCCTTCT 486
QY 645 TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704
DB 487 TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546
QY 705 TCAGTTGTTGTTTTCAGCAGTACTAAGGATGTTGAAGATGCCACCTCTCGACATGCTTAAA 764
DB 547 TCAGTTGTTGTTTTCAGCAGTACTAAGGATGTTGAAGATGCCACCTCTCGACATGCTTAAA 606
QY 765 GGAATAGCAGACAGACAGTATGGAGGGGTACTTGCACAAACAGATAAAGAGTCCCTGGA 824
DB 607 GGAATAGCAGACAGACAGTATGGAGGGGTACTTGCACAAACAGATAAAGAGTCCCTGGA 566
QY 825 TTAGTGCTAATCAAGATTGCGCTTTCCTGAGTGGATTTCACCAACATTCAGAGAAACA 884
DB 667 TTAGTGCTAATCAAGATTGCGCTTTCCTGAGTGGATTTCACCAACATTCAGAGAAACA 726
QY 885 AATCAACTAAAACAAAATGCGAGAAATCTTCTCTAAAGTCAAACTGTGAGTTG 944
DB 727 AATCAACTAAAACAAAATGCGAGAAATCTTCTCTAAAGTCAAACTGTGAGTTG 786
QY 945 TATTCAGAGCCAGTGTAGCGCTTCTTAAACACTGAATAACCCCAAGGATTTTCAAGAAATTG 1004
DB 787 TATTCAGAGCCAGTGTAGCGCTTCTTAAACACTGAATAACCCCAAGGATTTTCAAGAAATTG 846
QY 1005 AATAACCAACTAAGAAACATGACCATTTGATGAAAGAACTGACCAATAGTCTCTGCA 1064
DB 847 AATAACCAACTAAGAAACATGACCATTTGATGAAAGAACTGACCAATAGTCTCTGCA 906
QY 1065 TATTTATTTGGATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 1124
DB 907 TATTTATTTGGATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 966
QY 1125 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTCGGATCTATCGAGAGGCATGA 1184
DB 967 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTCGGATCTATCGAGAGGCATGA 1026
QY 1185 CCATGGGTTCCGAATATTTTCAATGTCACCAACGGGCAGATTCCTCTCGGTGACACCTT 1244
DB 1027 CCATGGGTTCCGAATATTTTCAATGTCACCAACGGGCAGATTCCTCTCGGTGACACCTT 1086
QY 1245 GACAACTCTCGAGTGACAATAGTAAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1304
DB 1087 GACAACTCTCGAGTGACAATAGTAAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1146
QY 1305 CCTACCTTTAGTTCACCTGCTATTTGAAGTCACTTCATCGCATCGAAGGGCTGTCCTCAG 1364
DB 1147 CCTACCTTTAGTTCACCTGCTATTTGAAGTCACTTCATCGCATCGAAGGGCTGTCCTCAG 1206
QY 1365 AAGTTTATTTACCTAAATGATGATGTTGTTGGAAGGATGTTCTGGCCAGATGATTTT 1424
DB 1207 AAGTTTATTTACCTAAATGATGATGTTGTTGGAAGGATGTTCTGGCCAGATGATTTT 1266
QY 1425 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCGCAG 1484
DB 1267 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCGCAG 1326
QY 1485 GGTGTCGCCAGGTTCTCGGATTAAGGATGCTTATTTGCAAGGCTTGTAAATTAATTCAGCC 1544
DB 1327 GGTGTCGCCAGGTTCTCGGATTAAGGATGCTTATTTGCAAGGCTTGTAAATTAATTCAGCC 1386
QY 1545 TGCAGATTGGGATGTTGGGATTTGCTCTTGAAAACAGTGGAGGAGGTGCTTATATTGAGGA 1604
DB 1387 TGCAGATTGGGATGTTGGGATTTGCTCTTGAAAACAGTGGAGGAGGTGCTTATATTGAGGA 1446
QY 1605 GGTGGAGGTACTGGGAGTATGAGTGTGACACCCCTGACAGTTGGTGGAGGATTAAC 1664
DB 1447 GGTGGAGGTACTGGGAGTATGAGTGTGACACCCCTGACAGTTGGTGGAGGATTAAC 1506
QY 1665 AGTGTCTTACTGTAATCAGGAGTGTGCGAAATTCCTGGCTCGCTGATAAGTCTTGTGAC 1724

Db 1507 AGTGCTCTTACTGTAATCAGGAGTGTGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1566
Qy 1725 CAAGCATGCAATGCTTTGCTGCTGTGGGTTTGATGCTGGCGACTGTGGGCGAAGATCAATTTT 1784
Db 1567 CAAGCATGCAATGCTTTGCTGCTGTGGGTTTGATGCTGGCGACTGTGGGCGAAGATCAATTTT 1626
Qy 1785 CATGAATGTATAAAGTATGCTCTCTCCCAACAGAGTCACTATATATTTTCCAAAAGGT 1844
Db 1627 CATGAATGTATAAAGTATGCTCTCTCCCAACAGAGTCACTATATATTTTCCAAAAGGT 1686
Qy 1845 GAATGCTGCTTATTTTCAAGTTTGCAGAGTAGCCAAAAGAGAGTGTGAAGTGTGCTTAT 1904
Db 1687 GAATGCTGCTTATTTTCAAGTTTGCAGAGTAGCCAAAAGAGAGTGTGAAGTGTGCTTAT 1746
Qy 1905 AGTGACATCCAAATATTCGACATGCTTTCTATTGCCCAAGTGGAAAACCATCCACCTC 1964
Db 1747 AGTGACATCCAAATATTCGACATGCTTTCTATTGCCCAAGTGGAAAACCATCCACCTC 1806
Qy 1965 ATAATGCACAGTGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACA 2024
Db 1807 ATAATGCACAGTGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACA 1866
Qy 2025 AACCATGAGAGTTCAAAATGCAGATTAACAGTGGNGTGACACACAGGAGGAGGACCAAA 2084
Db 1867 AACCATGAGAGTTCAAAATGCAGATTAACAGTGGNGTGACACACAGGAGGAGGACCAAA 1926
Qy 2085 CTGAATCTACGCGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAAACACTTTCTTCCA 2144
Db 1927 CTGAATCTACGCGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAAACACTTTCTTCCA 1586
Qy 2145 GAGCGGGAATCTTTTGTAGGATTAATCCCAAGAAAACGCTTCCCGAAGTTTAAGAGA 2204
Db 1987 GAGCGGGAATCTTTTGTAGGATTAATCCCAAGAAAACGCTTCCCGAAGTTTAAGAGA 2046
Qy 2205 CATGATGTAACTCAACAGAGAGAGCCAGAGAGAGTGAATAATTCCTCGTTAAATATT 2264
Db 2047 CATGATGTAACTCAACAGAGAGAGCCAGAGAGAGTGAATAATTCCTCGTTAAATATT 2106
Qy 2265 TCACCTCTTCCAAAGAGCCAGCTTGAGTCTCAATACCTTGGATTGCAACTGGAAACAT 2324
Db 2107 TCACCTCTTCCAAAGAGCCAGCTTGAGTCTCAATACCTTGGATTGCAACTGGAAACAT 2166
Qy 2325 GGAGACATCACTTTGAAGGATCAATTTGTCGAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2384
Db 2167 GGAGACATCACTTTGAAGGATCAATTTGTCGAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2226
Qy 2385 ATGAACCTCACAGCATGCTTAAATATAAATAAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2444
Db 2227 ATGAACCTCACAGCATGCTTAAATATAAATAAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2286
Qy 2445 AGTTTGTGCTCCACAGGAAAACAGGTTCAATAAAGCATCTGCCAAGAGCTTAGGA 2504
Db 2287 AGTTTGTGCTCCACAGGAAAACAGGTTCAATAAAGCATCTGCCAAGAGCTTAGGA 2346
Qy 2505 GTGTCTGAAAGATTGCGAGAGTTGACATTTTCTCGAGTGAAGTGAATAAGTGAATGTGAT 2564
Db 2347 GTGTCTGAAAGATTGCGAGAGTTGACATTTTCTCGAGTGAAGTGAATAAGTGAATGTGAT 2406
Qy 2565 GACAGGGTGCAGATCCACCCCTGGATTGGAGACCAAGAGAGATTTAGAGTGGAACT 2624
Db 2407 GACAGGGTGCAGATCCACCCCTGGATTGGAGACCAAGAGAGATTTAGAGTGGAACT 2466
Qy 2625 CACACCCAAAACCATAGGCGGAAATGTGCAAAAGAAAAGCCGCCATCTCTGATTGTT 2684
Db 2467 CACACCCAAAACCATAGGCGGAAATGTGCAAAAGAAAAGCCGCCATCTCTGATTGTT 2526
Qy 2685 CCACCTGGAAGCCAGATGACAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGT 2744
Db 2527 CCACCTGGAAGCCAGATGACAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGT 2586
Qy 2745 AGRATGGAGAAATGCTGAAATACATAGCCGCTTACTGAAGTGTACTTGAAGAAAG 2804
Db 2587 AGRATGGAGAAATGCTGAAATACATAGCCGCTTACTGAAGTGTACTTGAAGAAAG 2646

Qy 2805 CTGCAGCAATTACAGATAGTTACTTGGCTTTTGGCCATGGGAGAAAAAAGTATTTTC 2864
Db 2647 CTGCAGCAATTACAGATAGTTACTTGGCTTTTGGCCATGGGAGAAAAAAGTATTTTC 2706
Qy 2865 CAAGATCTTCTCGACGAAGAAGAGTCAATTGAAGACAAATTTGGCATACTTCACTGATAGC 2924
Db 2707 CTAGATCTTCTCGACGAAGAAGAGTCAATTGAAGACAAATTTGGCATACTTCACTGATAGC 2766
Qy 2925 AAAATATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCCAGATATGTAATAAA 2984
Db 2767 AAGATATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCCAGATATGTAATAAA 2826
Qy 2985 ATTCTAAATAGCAAGTTTGGATTCAATCGCGGAAGTCCCTGCATCGCTCAATG 3044
Db 2827 ATTCTAAATAGCAAGTTTGGATTCAATCGCGGAAGTCCCTGCATCGCTCAATG 2886
Qy 3045 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGC 3104
Db 2887 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGC 2946
Qy 3105 TCATTTTCAAAAGTGGCCATTCTGAGGATATGAGATATGCTTCTCTTATTTTATAT 3164
Db 2947 TCATTTTCAAAAGTGGCCATTCTGAGGATATGAGATATGCTTCTCTTATTTTATAT 3006
Qy 3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224
Db 3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066
Qy 3225 CAATCTGTGCTGTCTGTCAGAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3284
Db 3067 CAATCTGTGCTGTCTGTCAGAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3126
Qy 3285 CCGTTAAGTTTGCAGGATTTGACAGGCTTGAAACACATGCTTAATAAATTTGCTCAAAAATG 3344
Db 3127 CCGTTAAGTTTGCAGGATTTGACAGGCTTGAAACACATGCTTAATAAATTTGCTCAAAAATG 3186
Qy 3345 CTTCTGCTGTATACACAGCACTAAATAATTCACCAACTCAGGAATCTCTACTATGAT 3404
Db 3187 CTTCTGCTGTATACACAGCACTAAATAATTCACCAACTCAGGAATCTCTACTATGAT 3246
Qy 3405 CCGAACCTGCCACCGGTCACTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAA 3464
Db 3247 CCGAACCTGCCACCGGTCACTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAA 3306
Qy 3465 ATCCACAAAGCATATAAGGACAAAACAAATATAGTTTGAATCATGGGAGAGAGAA 3524
Db 3307 ATCCACAAAGCATATAAGGACAAAACAAATATAGTTTGAATCATGGGAGAGAGAA 3366
Qy 3525 ATCGCTTTTAAATGATTTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3584
Db 3367 ATCGCTTTTAAATGATTTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3426
Qy 3585 AGAAAAACCTTAGGAAGTTTGTTCCTGAATGACAACTTGACCAATTCATAAAGAT 3644
Db 3427 AGAAAAACCTTAGGAAGTTTGTTCCTGAATGACAACTTGACCAATTCATAAAGAT 3486
Qy 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGACATCTTATGAATCCATGTTCCCATACCTTCC 3704
Db 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGACATCTTATGAATCCATGTTCCCATACCTTCC 3546
Qy 3705 CAATTTGAACCTGCCAAGAGATTCGAAAACCGTTTCTTCATATGCAATGAGCTGAGGAA 3764
Db 3547 CAATTTGAACCTGCCAAGAGATTCGAAAACCGTTTCTTCATATGCAATGAGCTGAGGAA 3606
Qy 3765 TGGAGGGCTTATCGACAAATTTGAGTTTGGCCCATCTGTACTAGCAACATTTGAT 3824
Db 3607 TGGAGGGCTTATCGACAAATTTGAGTTTGGCCCATCTGTACTAGCAACATTTGAT 3666
Qy 3825 ATGTTTACTATATCTCAATTTTTTGGCTGAGCAGTTTAATTTGCACTTAAGCGGAAGATTT 3884
Db 3667 ATGTTTACTATATCTCAATTTTTTGGCTGAGCAGTTTAATTTGCACTTAAGCGGAAGATTT 3726

1375	Db		GGTGGGAA	TGCTCTGGAACAGTGGAGGAGTCGCTATATTGCAGAGAGTGGAGGTACT	1434
1617	Qy	GGGAGTAT	TGGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1676	
1435	Db	GGGAGTAT	TGGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1494	
1677	Qy	TGTAATCAGG	ATGTGCGAATTCTGTGCTCGCTGATTAAGTCTGTGTGACCAAGCATGCAAT	1736	
1495	Db	TGTAAATCAGG	ATGTGCGAATTCCTGGCTCGCTGATAAAGTCTGTGTGACCAAGCATGCAAT	1554	
1737	Qy	GTCTTGCTCTGTGG	TTTGATGCTGGCGACTGTGGCGAAGATCATTTTTCATGAATTCGTAT	1796	
1555	Db	GTCTTGCTCTGTGG	TTTGATGCTGGCGACTGTGGCGAAGATCATTTTTCATGAATTCGTAT	1614	
1797	Qy	AAAGTGTATCCTT	CTCCAAACACAGACTCACTATATTATTCCAAAGGTGAATGCTCGCT	1856	
1615	Db	AAAGTGTATCCTT	CTCCAAACACAGACTCACTATATTATTCCAAAGGTGAATGCTCGCT	1674	
1857	Qy	TATTTACGCTTTG	CAGAAGTAGCCAAAGAGGAGTTGAAGTGCTATAGTGACAATCCA	1916	
1675	Db	TATTTACGCTTTG	CAGAAGTAGCCAAAGAGGAGTTGAAGTGCTATAGTGACAATCCA	1734	
1917	Qy	ATAATTCGACATG	CTCTATTGGCAAAGTGGAAAAACCATCCACCTTCATAATGCACAGT	1976	
1735	Db	ATAATTCGACATG	CTCTATTGGCAAAGTGGAAAAACCATCCACCTTCATAATGCACAGT	1794	
1977	Qy	GGATGAATGCCA	CCAACAATTAATCTACGTTTCCAAATACAAAACGATGAAGAG	2036	
1795	Db	GGATGAATGCCA	CCAACAATTAATCTACGTTTCCAAATACAAAACGATGAAGAG	1854	
2037	Qy	TTCAAATTCAGAT	AAACAGTGGAGTGGACACAAGGGAGGGACCAAACTTGAATCTTACG	2096	
1855	Db	TTCAAATTCAGAT	AAACAGTGGAGTGGACACAAGGGAGGGACCAAACTTGAATCTTACG	1914	
2097	Qy	GCCCAGAGGGT	TACGAAAAATTTAGTTAGTCCCAATAACCTTCTCCAGAGGGCGAAATC	2156	
1915	Db	GCCCAGAGGGT	TACGAAAAATTTAGTTAGTCCCAATAACCTTCTCCAGAGGGCGAAATC	1974	
2157	Qy	CTTTTTCAGAGAT	TATCCCAAAGAAAAACGCTTCCGAAAGTTTAAGGACATGATGTTAAC	2216	
1975	Db	CTTTTTCAGAGAT	TATCCCAAAGAAAAACGCTTCCGAAAGTTTAAGGACATGATGTTAAC	2034	
2217	Qy	TCAACAGGAGAG	CGCCAGGAAGAGTGCAAAATTCCTCGTGAATATTTCCACTCCTTCCA	2276	
2035	Db	TCAACAGGAGAG	CGCCAGGAAGAGTGCAAAATTCCTCGTGAATATTTCCACTCCTTCCA	2094	
2277	Qy	AAAGACCCCAAGT	GTAGTCTCAATACTTGGAATTTGCAATGGAAACATGGAGACATCACT	2336	
2095	Db	AAAGACCCCAAGT	GTAGTCTCAATACTTGGAATTTGCAATGGAAACATGGAGACATCACT	2154	
2337	Qy	TTGAAAGGATAC	AAATTTGTCACAGTCAGCTCTGTGAGATCATTTCTGTGATGAATCCACAG	2396	
2155	Db	TTGAAAGGATAC	AAATTTGTCACAGTCAGCTCTGTGAGATCATTTCTGTGATGAATCCACAG	2214	
2397	Qy	CATGCTAAAAAT	AAAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT	2456	
2215	Db	CATGCTAAAAAT	AAAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT	2274	
2457	Qy	CCACAGAAAAA	CHAGTTTCAATAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAGA	2516	
2275	Db	CCACAGAAAAA	CHAGTTTCAATAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAGA	2334	
2517	Qy	TTGCAGAGGTTG	ACTTTTCTCTGAGTGAATGTAAGTGAATGGTCATGACCAAGGGTCAG	2576	
2335	Db	TTGCAGAGGTTG	ACTTTTCTCTGAGTGAATGTAAGTGAATGGTCATGACCAAGGGTCAG	2394	
2577	Qy	AATCCACCCCTG	CACTTGGAGACACAGCAAGATTTTAGATGGAAACTCAACCCAAAAA	2636	
2395	Db	AATCCACCCCTG	CACTTGGAGACACAGCAAGATTTTAGATGGAAACTCAACCCAAAAA	2454	
2637	Qy	ACCATAGCGG	GAATGTGACAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC	2696	

2455 ACATAGGCGGAAATGTGACAAAAAGAAAGCCCCCAATCTCTGATTGTTCTCCACTCGGAAGC 2514

2697 CAGATGACAAAAAGAAAGAAATCACAGGGAAGAAAAAGAACAGTAGAATGGAGAA 2756

2515 CAGATGACAAAAAGAAAGAAATCACAGGGAAGAAAAAGAACAGTAGAATGGAGAA 2574

2757 AATGCTGAAATTCATGAGCGGTACTGGAAGTGTACTTTGGAAGAAAGCTCGACATTC 2816

2575 AATGCTGAAATTCATGAGCGGTACTGGAAGTGTACTTTGGAAGAAAGCTCGACATTC 2634

2817 ACAGATAGTACTCTGGCGTTTTCGCCATGGCAGAAAAAAAGTATTTCAGAGATCTCTC 2876

2635 ACAGATAGTACTCTGGCGTTTTCGCCATGGCAGAAAAAAAGTATTTCAGATCTCTC 2694

2877 GACGAAGAAGATCAATGAAGACAAATGGCATACTTCAGTAGCAAAAAATCTCGG 2936

2695 GAGGAAGAAGATCAATGAAGACAAATGGCATACTTCAGTAGCAAGAATCTCGG 2754

2937 AGGCAACTAAAAGATACATTTGCAGATTCCTCCAGATATGTAATAAAATTTCTAAATAGC 2996

2755 AGGCAACTAAAAGATACATTTGCAGATTCCTCCAGATATGTAATAAAATTTCTAAATAGC 2814

2997 AAGTTTGGATTCCATATCGCGAAAGTCCCTGCTCCATCGCTCACATGATTCACCGGATT 3056

2815 AAGTTTGGATTCCATATCGCGAAAGTCCCTGCTCCATCGCTCACATGATTCACCGGATT 2874

3057 GTTATCGAAGATCTCAGATATGTTCCCTGGAAGAAATTGACAGACGTCATTTACAAA 3116

2875 GTTATCGAAGATCTCAGATATGTTCCCTGGAAGAAATTGACAGACGTCATTTACAAA 2934

3117 GTCGGCATTCTGAGGATATGCAAGTTGGCTTCTCTATTATTTATTTCTCATGAGTGA 3176

2935 GTCGGCATTCTGAGGATATGCAAGTTGGCTTCTCTATTATTTATTTCTCATGAGTGA 2994

3177 GTCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGHTC 3236

2995 GTCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGHTC 3054

3237 TTGCTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACGTCGGTTAAGTTG 3296

3055 TTGCTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACGTCGGTTAAGTTG 3114

3297 CAGGATTTGACAGGCTGCGAACACATGCTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT 3356

3115 CAGGATTTGACAGGCTGCGAACACATGCTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT 3174

3357 ATCAGCAGCTAAATATATTTACCAACCTCAGGAATCTCTATGATCCCAACCTGCGCA 3416

3175 ATCAGCAGCTAAATATATTTCCACCACCTCAGGAATCTCTATGATCCCAACCTGCGCA 3234

3417 CCGGTCACTAAAAGTCTAGTAAACAACTGTAAACAGTAATTCGAAAAATCCCAAGACA 3476

3235 CCGGTCACTAAAAGTCTAGTAAACAACTGTAAACAGTAATTCGAAAAATCCCAAGACA 3294

3477 TATAAGGAACAAAAACAATATAGGTTTGAATCATGGGAAGAAAGAAATCGCTTTTAAA 3536

3295 TATAAGGAACAAAAACAATATAGGTTTGAATCATGGGAAGAAAGAAATCGCTTTTAAA 3354

3537 ATGATTCGTACCAAGCTTCTCATGTGGTTGGCCAGTTGCGATGACATAAGAAAAAACCT 3596

3355 ATGATTCGTACCAAGCTTCTCATGTGGTTGGCCAGTTGCGATGACATAAGAAAAAACCT 3414

3597 AGGAAGTTTGTGTGCTGAATGACCAATTCAGCCAACTATAAGATGCTCAGACAGTG 3656

3415 AGGAAGTTTGTGTGCTGAATGACCAATTCAGCCAACTATAAGATGCTCAGACAGTG 3474

3657 AAGGCTGTCTCAGGACTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGACTG 3716

3475 AAGGCTGTCTCAGGACTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGACTG 3534

3717 CCAAGAGATATCGAAAACCGTTTCTTCATATGCAATGAGCTGCAAGAAATGGAGGCTTAT 3776

3525 CCAAGAGATATCGAAAACCGTTTCTTCATATGCAATGAGCTGCAAGAAATGGAGGCTTAT 3594

Db 3523 CCAAGAGAGTATCGAAACCGTTTCTTCATATGATGATGCTGAGGAATGGAGGGCTTAT 3582
QY 3777 CGAGACAAATTGAAGT 3792
Db 3583 CGAGACAAATTGAAGT 3598
RESULT 10
AAD62490
ID AAD62490 standard; DNA; 3600 BP.
XX
AC AAD62490;
XX
DT 15-JAN-2004 (first entry)
XX
DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase cDNA.
XX
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KW gastrointestinal; gene; ds; lectin resistant cell; deoxymannojirimycin;
KW kifunensine; glycosylation inhibition.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..3600
FT /*tag= a
FT /product= "N-acetylglucosamine-1 (GlcNAc)-
FT phosphotransferase"
XX
FN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield WM;
XX
DR WPI; 2003-810984/76.
XX
DR P-PSDB; ABW01487.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 11; Page 8-10; 46pp; English.
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
CC (GlcNAc)-phosphotransferase cDNA
SQ Sequence 3600 BP; 1151 A; 736 C; 798 G; 915 T; 0 U; 0 Other;
Query Match 61.9%; Score 3465.6; DB 10; Length 3600;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 3477; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 297 AGCCGAGATCAATACATGTTTGTTCATTCCTATAGACATATTCCTGGAAGTCC 356
Db 103 AGCCGAGATCAATACATGTTTGTTCATTCCTATAGACATATTCCTGGAAGTCC 162

QY 357 TTTTCAAAATCGGCTTTGTTCTGCCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 416
Db 163 TTTTCAAAATCGGCTTTGTTCTGCCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 222
QY 417 ACAGATCTTTGAATCTACTGAAAGAACTTACAGCAGAGTTCAGAGAAACAGATGAGGAGGAGCAG 476
Db 223 ACAGATCTTTGAATCTACTGAAAGAACTTACAGCAGAGTTCAGAGAAACAGATGAGGAGGAGCAG 282
QY 477 AAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTTACTTAAGAGAGTGAAGAG 536
Db 283 AAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTTACTTAAGAGAGTGAAGAG 342
QY 537 CAGTTAGAGTGTGTTGCTAAACACATGCTATTAAAGGTGCCAATGCTTGTACTGACCCAGCC 596
Db 343 CAGTTAGAGTGTGTTGCTAAACACATGCTATTAAAGGTGCCAATGCTTGTACTGACCCAGCC 402
QY 597 CTGCCAGCAACATCACTGAGGAGCTGCTTATTCCTTATTCCTTCTTTTCTTTTCTTTTCTGCC 656
Db 403 CTGCCAGCAACATCACTGAGGAGCTGCTTATTCCTTATTCCTTCTTTTCTTTTCTTTTCTGCC 462
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCTTCTACCAATGCTCAGTTGTTGTT 716
Db 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCTTCTACCAATGCTCAGTTGTTGTT 522
QY 717 TTTTCAAGTACTAAGGATGTTGAAGATGCCACTCTGACTGCTTAAAGGAATAGCAGA 776
Db 523 TTTTCAAGTACTAAGGATGTTGAAGATGCCACTCTGACTGCTTAAAGGAATAGCAGA 582
QY 777 CAGACAGTATGAGGGGGTACTTGACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 836
Db 583 CAGACAGTATGAGGGGGTACTTGACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 642
QY 837 CAAAGTTTGGCTTTCTGAGTGGATTTCACCAACATTCAGAGAAACAAATCAACTAAAA 896
Db 643 CAAAGTTTGGCTTTCTGAGTGGATTTCACCAACATTCAGAGAAACAAATCAACTAAAA 702
QY 897 ACAAATTCGCAAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 956
Db 703 ACAAATTCGCAAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 762
QY 957 AGTGAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGCAAACT 1016
Db 763 AGTGAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGCAAACT 822
QY 1017 AAGAAGAACATGACCAATTCATGAAAGAACTGAACATTAAGTCTGTCATATTTATG 1076
Db 823 AAGAAGAACATGACCAATTCATGAAAGAACTGAACATTAAGTCTGTCATATTTATG 882
QY 1077 GATCTGAGCGCATCAGCCAGTCTAGCAGGATGAGACATCTCTGCCAGTGGTTTGA 1136
Db 883 GATCTGAGCGCATCAGCCAGTCTAGCAGGATGAGACATCTCTGCCAGTGGTTTGA 942
QY 1137 GATAACGAAAGTCTGAGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCG 1196
Db 943 GATAACGAAAGTCTGAGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCG 1002
QY 1197 AATATTTTCAATGTCACCAACGGGAGATTCATCTGCTGTAACCTTGACATCTCGA 1256
Db 1003 AATATTTTCAATGTCACCAACGGGAGATTCATCTGCTGTAACCTTGACATCTCGA 1062
QY 1257 GTCACAATATGACACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTAGT 1316
Db 1063 GTCACAATATGACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTAGT 1122
QY 1317 TCACCTGCTATTGAAAGTCAATTCATTCGATCGAAGGGCTGTCCAGAGTTTATTAC 1376
Db 1123 TCACCTGCTATTGAAAGTCAATTCATTCGATCGAAGGGCTGTCCAGAGTTTATTAC 1182
QY 1377 CTAATATGATGATGCTATGTTTGGAGAGGATGCTGCCAGATGATTTTTACAGTCACTCC 1436
Db 1183 CTAATATGATGATGCTATGTTTGGAGAGGATGCTGCCAGATGATTTTTACAGTCACTCC 1242
QY 1437 AAAGGCCAGAGGTTTATTTTGACATGGCCTGTGCCAAACTGTGCCGAGGGCTGCCAGGT 1496


```
QY 3657 AAGGCTGTTCTCAGGACTTCTATGAATCGATGTTCCCATACCTTCCCAATTGAACTG 3716
DB 3463 AAGGCTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTGAACTG 3522
QY 3717 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGATGCTGACGAGGATGAGGCGTTAT 3776
DB 3523 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGATGCTGACGAGGATGAGGCGTTAT 3582
QY 3777 CGAGACAAATTGAAAGT 3792
DB 3583 CGAGACAAATTGAAAGT 3598

RESULT 11
AD62649
ID AD62649 standard; cDNA; 3600 BP.
XX
AC AD62649;
XX
DT 15-JAN-2004 (first entry)
XX
DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase cDNA.
XX
KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
KW N-acetylglucosamine-1-phosphotransferase; gene; gene therapy; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..3600
FT /tag= a
FT /product= "N-acetylglucosamine-1 (GlcNAc)-
FT phosphotransferase"
XX
PN US2003124653-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023890.
XX
PR 21-DEC-2001; 2001US-00023890.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield WM;
XX
WPI; 2003-810985/76.
DR P-PSDB; ABW01536.
XX
PT Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX
PS Claim 11; Page 8-10; 46pp; English.
XX
CC The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
CC phosphotransferase cDNA
XX
SQ Sequence 3600 BP; 1151 A; 736 C; 798 G; 915 T; 0 U; 0 Other;
Query Match 61.9%; Score 3465.6; DB 10; Length 3600;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 3477; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 297 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATGCTGGAAGTCC 356
DB 103 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATGCTGGAAGTCC 162
```

```
QY 357 TTTGAGATCGGTTTGTCTGCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 416
DB 163 TTTGAGATCGGTTTGTCTGCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 222
QY 417 ACAGATCTTTGAATCTACTGAAGAACTACAGCAGGTTCAGAGAAACAGATGAGGAGGACAG 476
DB 223 ACAGATCTTTGAATCTACTGAAGAACTACAGCAGGTTCAGAGAAACAGATGAGGAGGACAG 282
QY 477 AAGCAATGAGAGAAATCCTTGGGAAAAACACAAAGGAACCTTAAAGAGAGTGAAGAG 536
DB 283 AAGCAATGAGAGAAATCCTTGGGAAAAACACAAAGGAACCTTAAAGAGAGTGAAGAG 342
QY 537 CAGTTAGAGTGTGTTGCTAACACACTGCTTAAAGGTGCCAATCTTGTACTGACCCAGCC 596
DB 343 CAGTTAGAGTGTGTTGCTAACACACTGCTTAAAGGTGCCAATCTTGTACTGACCCAGCC 402
QY 597 CTGCCAGCAACATCACCCTGAAGGACCTGCCATCTCTTTATCTCTTTTCACTTGTGCC 656
DB 403 CTGCCAGCAACATCACCCTGAAGGACCTGCCATCTCTTTATCTCTTTTCACTTGTGCC 462
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGTCTCAGTTGTGTT 716
DB 463 AGTGACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGTCTCAGTTGTGTT 522
QY 717 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 776
DB 523 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 582
QY 777 CAGACAGTATGAGGGGTACTTGCACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 836
DB 583 CAGACAGTATGAGGGGTACTTGCACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 642
QY 837 CAAGATTTCGCTTTCCTGAGTGGATTCCACCAACATTCAGAGAAACAAATCACTAAAA 896
DB 643 CAAGATTTCGCTTTCCTGAGTGGATTCCACCAACATTCAGAGAAACAAATCACTAAAA 702
QY 897 ACAAAATGCCAGAAATCTTTCCTCTAAAGTCAAACTGTTCGAGTTGTTATCAGAGGCC 956
DB 703 ACAAAATGCCAGAAATCTTTCCTCTAAAGTCAAACTGTTCGAGTTGTTATCAGAGGCC 762
QY 957 AGTGTAGCGCTTCTAAACTGAATACCCCAAGGATTTTCAAGATTGTAATGAAGAAACT 1016
DB 763 AGTGTAGCGCTTCTAAACTGAATACCCCAAGGATTTTCAAGATTGTAATGAAGAAACT 822
QY 1017 AAGAGAACATGACCATTTGATGGAAGAACTGACCATAGTCTCTGCATATTTATTATGG 1076
DB 823 AAGAGAACATGACCATTTGATGGAAGAACTGACCATAGTCTCTGCATATTTATTATGG 882
QY 1077 GATCTGAGCGCATGAGCCAGTCTAAGAGAGTGAAGACATCTCTGCCAGTCTGTTTGA 1136
DB 883 GATCTGAGCGCATGAGCCAGTCTAAGAGAGTGAAGACATCTCTGCCAGTCTGTTTGA 942
QY 1137 GATAACCAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCGG 1196
DB 943 GATAACCAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCGG 1002
QY 1197 AATATTTTCATTTGTCACCAACCGGCGAGTTCATCTCTGCTGAACTTGCATCTCGA 1256
DB 1003 AATATTTTCATTTGTCACCAACCGGCGAGTTCATCTCTGCTGAACTTGCATCTCGA 1062
QY 1257 GTGACAAATAGTAACACACAGGATGTTTTCGAAATTTTGAGCCACTTGCCTACCTTTAGT 1316
DB 1063 GTGACAAATAGTAACACACAGGATGTTTTCGAAATTTTGAGCCACTTGCCTACCTTTAGT 1122
QY 1317 TCACCTGCTATTGAAAGTCAATTCATTCGATCGAAGGCTGTCCAGAGTTTATTATAC 1376
DB 1123 TCACCTGCTATTGAAAGTCAATTCATTCGATCGAAGGCTGTCCAGAGTTTATTATAC 1182
QY 1377 CTAATGATGATGTCATGTTTGGGAGGATGTCGCCAGATGATTTTACAGTCACATCC 1436
DB 1183 CTAATGATGATGTCATGTTTGGGAGGATGTCGCCAGATGATTTTACAGTCACATCC 1242
```

QY	1437	AAAGCCAGAGGTTTATTTGACATGGCTGTGCCAACTGTGCGGAGGCTGCCAGGT	1496	DB	2323	TTGCGAGGTTGACCTTTCTGCGAGTGATGTAAGTGATGTCATGACGAGGTCAG	2382
DB	1443	AAAGCCAGAGGTTTATTTGACATGGCTGTGCCAACTGTGCGGAGGCTGCCAGGT	1502	QY	2577	AATCCACCCCTGGACTTGGAGACCAAGAGATTTAGAGTGGAACTCACACCCAAAA	2636
QY	1497	TCCTGGATTAAAGGATGGCTATTTGACAAAGCTTGTAAATTAATCAGCCTCGGATGGAT	1556	DB	2383	AATCCACCCCTGGACTTGGAGACCAAGAGATTTAGAGTGGAACTCACACCCAAAA	2442
DB	1503	TCCTGGATTAAAGGATGGCTATTTGACAAAGCTTGTAAATTAATCAGCCTCGGATGGAT	1562	QY	2637	ACCATAGCGGAAATGTCACAAAGAAAGAGCCCATCTCTGATTTGTTCCACTGGAAGC	2696
QY	1557	GGTGGGATTCCTCTGGAAACAGTGGAGGAGTGCCTATATTTGACAGAGTGGAGTACT	1616	DB	2443	ACCATAGCGGAAATGTCACAAAGAAAGAGCCCATCTCTGATTTGTTCCACTGGAAGC	2502
DB	1563	GGTGGGATTCCTCTGGAAACAGTGGAGGAGTGCCTATATTTGACAGAGTGGAGTACT	1422	QY	2697	CAGATGACAAAGAAAGAAATCAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2756
QY	1617	GGGAGATTGAGTGGACACCCCTGGCAGTTTGGTGGAGAAATAAACAGTGTCTTTAC	1676	DB	2503	CAGATGACAAAGAAAGAAATCAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2562
DB	1423	GGGAGATTGAGTGGACACCCCTGGCAGTTTGGTGGAGAAATAAACAGTGTCTTTAC	1482	QY	2757	AATGCTGAATCATACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGATTAC	2816
QY	1677	TGTAATCAGGATGTGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT	1736	DB	2563	AATGCTGAATCATACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGATTAC	2622
DB	1483	TGTAATCAGGATGTGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT	1542	QY	2817	ACAGATAGTTACTTGGGCTTTTGGCATGGAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2876
QY	1737	GTCTTGCTCTGGGTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT	1796	DB	2623	ACAGATAGTTACTTGGGCTTTTGGCATGGAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2682
DB	1543	GTCTTGCTCTGGGTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT	1502	QY	2877	GACGAAGAGAGTCTTGAAGACACAAATTTGGCATACTTCTGATAGCAAAATTTACTGGG	2936
QY	1797	AAAGTATTCCTTCCCAACAGACTCACTATATTTTCCAAAGGTGAATGCCCTGCCT	1856	DB	2683	GACGAAGAGAGTCTTGAAGACACAAATTTGGCATACTTCTGATAGCAAAATTTACTGGG	2742
DB	1603	AAAGTATTCCTTCCCAACAGACTCACTATATTTTCCAAAGGTGAATGCCCTGCCT	1662	QY	2937	AGGCAACTTAAAGATACATTTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC	2996
QY	1857	TATTTGAGTTTGCAGAGTAGCCAAAGAGGAGTTCAAGGTGCCTATAGTGAATATCCA	1916	DB	2743	AGATCAAGAGAGATACATTTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC	2802
DB	1663	TATTTGAGTTTGCAGAGTAGCCAAAGAGGAGTTGAAGGTGCCTATAGTGAATATCCA	1722	QY	2997	AAGTTTGGATTCACATCCGGAAGTCCCTGCTCACATGCTCATGATTTGACCGGAT	3056
QY	1917	ATAATTCGACATGCTTATTTGCCAAAGTGGAAACCATCCACCTCATATGTCACAGT	1976	DB	2803	AAGTTTGGATTCACATCCGGAAGTCCCTGCTCACATGCTCATGATTTGACCGGAT	2862
DB	1723	ATAATTCGACATGCTTATTTGCCAAAGTGGAAACCATCCACCTCATATGTCACAGT	1782	QY	3057	GTATGCGAGAGTGCAGATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTACAAA	3116
QY	1977	GGATGAATGCCACCAATACATTTTAACTACGTTTCAAAATACAAACGATGAAGAG	2036	DB	2863	GTATGCGAGAGTGCAGATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTACAAA	2922
DB	1783	GGATGAATGCCACCAATACATTTTAACTACGTTTCAAAATACAAACGATGAAGAG	1842	QY	3117	GTGCGCCATTCGAGGATGATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGA	3176
QY	2037	TTCAAAATGACAGATAACAGTGGAGTGGACACAGGAGGAGGACCAAACTGAAATTTACG	2096	DB	2923	GTGCGCCATTCGAGGATGATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGA	2982
DB	1843	TTCAAAATGACAGATAACAGTGGAGTGGACACAGGAGGAGGACCAAACTGAAATTTACG	1902	QY	3177	GTGCGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCATCTGCTGTC	3236
QY	2097	GCCAGAGGTTTACGAAATTTAGTTAGTCCCAATACATCTTCTCCAGAGCGGAAATC	2156	DB	2983	GTGCGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCATCTGCTGTC	3042
DB	1903	GCCAGAGGTTTACGAAATTTAGTTAGTCCCAATACATCTTCTCCAGAGCGGAAATC	1962	QY	3237	TTGCTGACAGAGAAATCCGAAACACTGGCTACCAAGAAATTCACGAATTCGCCGTTAAGTTG	3296
QY	2157	CTTTTGGAGGATATCCCAAGAAACCGCTTCCGAGTTTAAAGACATGATGTTAAC	2216	DB	3043	TTGCTGACAGAGAAATCCGAAACACTGGCTACCAAGAAATTCACGAATTCGCCGTTAAGTTG	3102
DB	1963	CTTTTGGAGGATATCCCAAGAAACCGCTTCCGAGTTTAAAGACATGATGTTAAC	2022	QY	3297	CAGGATTTGACAGGTTCTGGAAACATGCTATAAATTTGCTCAAAATTTGCTTCTGCTGAT	3356
QY	2217	TCAACAGGAGAGCCAGGAGAGGTGAAATTTCCCTGGTAAATTTTCACTCTTCCCA	2276	DB	3103	CAGGATTTGACAGGTTCTGGAAACATGCTATAAATTTGCTCAAAATTTGCTTCTGCTGAT	3162
DB	2023	TCAACAGGAGAGCCAGGAGAGGTGAAATTTCCCTGGTAAATTTTCACTCTTCCCA	2082	QY	3357	ATCAGCGAGCTAAATATATTTCCCAACTCAGGAATCTCTACTATGATCCCAACTGCA	3416
QY	2277	AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAATGAGAGCATCACT	2336	DB	3163	ATCAGCGAGCTAAATATATTTCCCAACTCAGGAATCTCTACTATGATCCCAACTGCA	3222
DB	2083	AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAATGAGAGCATCACT	2142	QY	3417	CCGCTCACTAAAGCTTAGTAAACAACTGTAACCACTGAAGTGAAGTGAAGTGAAGTGA	3476
QY	2337	TTGAAAGGATACATTTTGTCCAGTTCAGCTTGTGATGATCAATTTCTGATGAATCAGAG	2396	DB	3223	CCGCTCACTAAAGCTTAGTAAACAACTGTAACCACTGAAGTGAAGTGAAGTGAAGTGA	3282
DB	2143	TTGAAAGGATACATTTTGTCCAGTTCAGCTTGTGATGATCAATTTCTGATGAATCAGAG	2202	QY	3477	TATAGGACAAACAAATATAGTTTCAATCATGGAAGAAAGAAATTCGCTTTTAAA	3536
QY	2397	CATCTTAAATATAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT	2456	DB	3283	TATAGGACAAACAAATATAGTTTCAATCATGGAAGAAAGAAATTCGCTTTTAAA	3342
DB	2203	CATCTTAAATATAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT	2262	QY	3537	ATGATTTCTGACCAAGTTTCTCATGTGTTGGCCAGTTTGGATGATGATGATGATGATGATG	3596
QY	2457	CCACAGGAAAAACAGTTTCAAAAGCATCTTGGCAAAACAGCTTAGGAGTGTCTGAAAGA	2516	DB	3343	ATGATTTCTGACCAAGTTTCTCATGTGTTGGCCAGTTTGGATGATGATGATGATGATGATG	3402
DB	2263	CCACAGGAAAAACAGTTTCAAAAGCATCTTGGCAAAACAGCTTAGGAGTGTCTGAAAGA	2322	QY	3597	AGGAAGTTTGTGCTGATGACAAACATGACCAATCATTAAGATGCTCAGACAGTG	3656

Db 1081 CCATCCTGGCTGAACCTTGACAACCCCTCGAGTGACCATAGTGACCCACAGGACATTTTC 1140
Qy 1287 CGAAATTTGAGCCACTGCTGCTACCTTGAAGTTCACCTGCTATTTGAAGTTCACATTCATCGC 1346
Db 1141 CAAAATCTGAGCCACTTGGCTTACTTTTCAGTTCCCTGCTATTGAAAGTTCATTTCAACCGC 1200
Qy 1347 ATCGAAGGGCTGCTCCAGAGTTTATTTACTAATCATGATGCTATGTTTGGCAAGGAT 1406
Db 1201 ATCGAAGGGCTGCTCCAGAGTTTATTTATTTCTAATGACGATGCTCATGTTTGGTAAGGAC 1260
Qy 1407 GTCTGGCCAGATGATTTTACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCT 1466
Db 1261 GTCTGGCCGAGGATTTTACAGCACTCCAAAGGCTCAAAGGTTTATTTGACATGGCCT 1320
Qy 1467 GTGCCAACTGTGCGAGGGTCCAGGTTCTCGGATTAAGGATGGCTATTTGACAAG 1526
Db 1321 GTGCCAACTGTGAGAGGGTCCCGGGCTCTCGGATAAGGACGGCTATTTGATAAG 1380
Qy 1527 GCTTGTAATTAATTCAGCTGCGATTTGGGATGGGGATTGCTCTGGAACAGTGGAGGG 1586
Db 1381 GCTTGTAATTAATTCAGCTGCGATTTGGGATGGGGAACTGCTCTGTAATTAATTCAGGG 1440
Qy 1587 AGTGGCTATATTGAGAGGGTGGAGTACTGGGAGTATTTGGAGTGGACACCCCTGGCAG 1646
Db 1441 AACCAGGTTTGTGCAAGAGGTGGGGTACCGGGATATTGGAGCTGGACAGCACTGGCAG 1500
Qy 1647 TTTGGTGAGGAATAAACAGTGTCTTACTGTATCAGGGATGTGCGAATTTCTGGCTC 1706
Db 1501 TTTGGTGAGGAATAAACAGTGTCTTACTGTATTAACCAAGGATGTGCAAACTCTGGCTG 1560
Qy 1707 GCTGATAAGTCTGTGACCAAGATGCAATGCTTTGCTGCTGGGTTTGAATGCTGGCAG 1766
Db 1561 GCTGACAGTCTGTGACCAAGCTGTAACTTTATCTCTGGGGTTTGAATGCTGGTGC 1620
Qy 1767 TGTGGCAGAGATCAATTTTATGAATTTATTAAGTATCTTCTCCAAACCAAGACTCAC 1826
Db 1621 TGTGACCAAGATCAATTTTATGAATTTATTAAGTATCACTTCTCCAAACCAAGACTCAC 1680
Qy 1827 TATATTTTCCAAAGGTGAATGCTGCTTATTTACAGCTTTGAGAAAGTAGCCAAAGA 1886
Db 1681 TATGTTGTCCCAAGGTGAATGCTGCTTATTTACAGCTTTGCAAAACATAGCCAGAAA 1740
Qy 1887 GAGTTGAGGTGCTATAGTGAATCCATATTTGACATGCTTCTATTTGCCAAG 1946
Db 1741 AGAATGAAGGAGCTTACAGGCAACCCCATCATCCGCCACGGCTCCATTTGCAACAAG 1800
Qy 1947 TGGAAACCATCCACCTCATATGACAGTGGAAATGAATGCGACACACATACATTTTAAAT 2006
Db 1801 TGGAAACCATCCACCTCATATGATGCGCGGGGATGAACGCCACCATATTTTAAAC 1860
Qy 2007 CTCAGTTTCAAAATACAAAGATGAAGTTTCAAAATGCGATTAACAGTGGAGGTGGAC 2066
Db 1861 CTCACCTTTCAAAAGCCCAAGAGTTCAGATCCAGATAGCAGTAGAGGTGGAC 1920
Qy 2067 ACAAGGAGGACCAAACTGAAATCTACGCGCCAGAGGGTTACGAAATTTAGTTAGT 2126
Db 1921 ACGAGGAGGCGCCAACTGAATTTTACACCCAGAGGCTATGAAGTTTGGTTAGC 1980
Qy 2127 CCCATAACACTTTTCCAGAGCGGAAATCTTTTGGAGATATTTCCAAAGAAAACGC 2186
Db 1981 CCAGTGACACCTTCTCTCAGGCTGACGCTGCTTTTGAAGATGTCCCAAGAGAAAACGC 2040
Qy 2187 TTCCCGAAGTTTAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAAA 2246
Db 2041 TTCCCAAGATCAGGAGCATGATGTAATTCGAACAGGAGATTTCCAGAGGAGGTGAAA 2100
Qy 2247 ATTCCTGTTAAATATTTTCACTCTTCCAAAGACGCCAGTGTAGTCTCAATACCTTG 2306
Db 2101 ATCCCCGGGTAAATATTTTCACTCTTCCAAAGAGGCCAGGTTGAGGCTGAGCAACTTG 2160
Qy 2307 GATTTGCAACTGGAACATGGAGACATCACTTTGAAGGATACAAATTTGTCAGTCAAGC 2366
Db 2161 GATTTGCAACTAGAACGTTGGAGACATCACTCTGAAAGGATATACTTGTCCAGTCAAGC 2220

Qy 2367 TTGCTGAGATCATTTCTGATGAACCTCACAGCATGCTAAAAATAAAAAATCAAGCTATATAA 2426
Db 2221 CTGCTTAAGGTCTTTCTGGGGAATTTCTAGATATAATAATAAACCTCAAGCT--AGG 2277
Qy 2427 ACAGATGAAAACAAATGACAGATTTGGTGGCTCCACAGGAAAACAGGTTTCATAAAGCATC 2486
Db 2278 ACCGATGAAAACAAAGGCAACCTGGAGGTCCACAGGAAAACCTTCTTCACAGACG--- 2333
Qy 2487 TTGCCAAACAGCTTAGGATGCTCTGAAGATTTGAGAGGTTGACTTTTCTGCAAGTCACT 2546
Db 2334 --TCCACATGGCTTTGCTGGTGAACACAGATCAGAGATGCACTGCCCCAGCAGACACA 2391
Qy 2547 GTAAAAGTGAATGTGTCATGACCCAGGGTCAGAAATCCACCCCTGGACTTTGGAGACCACAGCA 2606
Db 2392 GTGACCGTGAAGGCGGTGACCAAGCTTTGAATCCACCCCGGTGTTGGAGACCAATGCA 2451
Qy 2607 AGATTTAGAGTGGAAACTCACACCCAAAACCATAGCGGGAATGTGACAAAAGAAAAG 2666
Db 2452 AGATT-----GGCCCGAGCCTACACTAGCGGTGACTGTCTCCAAAAGAAC 2496
Qy 2667 CCCCCATCTCTGATTTGTTCCACTTGAAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGG 2726
Db 2497 CTTTCCACGCTGATGTTTCCCCCAGAAAGCCACTTGGCCAAAAGAAAGAGAGAGTGACAGG 2556
Qy 2727 AAAGAAAAGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGGCGTTACTGAA 2786
Db 2557 GCAGAA-----GGCAATGCTGTACCTGTAAAGAG 2586
Qy 2787 GTGTTTACTTGAAGAAAAGCTGACGATTTACACAGATAGTTACTTTGGGCTTTTGGCCATGG 2846
Db 2587 TTAGTCTGCGCAGACGGTTGCAGC-----AGAATTATCCAGGCTTTTGGCCCTGG 2637
Qy 2847 GAGAAAAAAGTATTTCCAAAGATCTTCTCGACGAGAAAGAGTCAATTTGAAGACACAATTG 2906
Db 2638 GAGAAAAAAGTATTTCCAAAGACCTTCTGATGAGGAAAGAGTCAATTTGAAGACCCAGTTG 2697
Qy 2907 GCATATCTTCACTGATGACAAAATACTGGGAGGCACTTAAAGATACATTTGCGAGATTC 2966
Db 2698 GCGTACTTTTACAGACCGCAACATACCGGAGGCACTTAAAGATACATTTGCGAGACTCC 2757
Qy 2967 CTCAGATATGTAATAAATAATTTCTAAATAGCAAGTTTGGATTCACATCCGCGAAAGTCCCT 3026
Db 2758 CTCAGATAGTCAATAAATAATTTCTCAACAGCAAGTTTGGATTCACATCCAGGAAAATGCCCT 2817
Qy 3027 GCTCATGCTCTACATGATTTGACCGGATTTGATGCAAGACTGCAAGATATGTTCCCT 3086
Db 2818 GCACACATGCCCAACATGATTTGACAGGATCGTTATGCAAGAACTCCAAAGATATGTTCCCT 2877
Qy 3087 GAAGAATTTGACACAGACGTCATTTCAAAAGTGCGCCATTTCTGAGGATATGCAAGTTGCC 3146
Db 2878 GAAGAATTTGACACAGCTTCATTTCAAGGTGCGTCACTCTGAGGACATGCAAGTTGCC 2937
Qy 3147 TTCTCTTATTTTATATCTCATGAGTGCAGTGCAGCCTGAATATATCTCAAGTCTTT 3206
Db 2938 TTCTCCTACTTTTATTAATCCTCATGAGTGCAGTTCAGCCCTCAATTAATTTCCCAAGTCTTT 2997
Qy 3207 GATGAGTTGTATACAGATCAATCTGCTGTCTGTCTGACAGAGAAATCCGAAACACTGGCT 3266
Db 2998 CATGAGTTAGACACAGACCAATCTGCTGTCTGTCTGATAGGAAATCCGAACTGTCGCC 3057
Qy 3267 ACCAGAAATTCAGAACTGCCGTTAAGTTTGAAGGATTTGACAGGTTGCGAAACATGCTA 3326
Db 3058 ACCAGAAATTCAGCACTTAAAGCTTGAAGGATTTGACAGGTTTGAAGACACATGTTA 3117
Qy 3327 ATAAATGCTCAAAATGCTTCTGCTGATACAGCAGCTAAATAATTTCCACCAACT 3386
Db 3118 ATAAATGCTCAAAATGCTTCCCGCTAATATCATCTCACTCAACACATCCACCGACT 3177
Qy 3387 CAGGAATCCTACTATGATCCCAACCTGCCACCGGTCATTAAGGTCATGTAACAACTGT 3446
Db 3178 CAGGAAGCATACTACGACCCCAACCTGCTCGGTCACTAAGAGTCTTGTCCACCAACTGT 3237

QY	3447	AAACCAAGTAACTGACAAAAATCCACAAAGCATATTAAGGACAAAAACAATAATAGTGTGGAA	3506
DB	3238	AAGCCAGTAACCTGACAAAGATCCACAAAGCCCTATTAAGACAAAGAACAAATACAGGTGGAA	3297
QY	3507	ATCATGGGAGAGAAGAAATCGCTTTTAAAAATGATTCGTACCAACGTTTCTCATGTGCTT	3566
DB	3298	ATCATGGAGAGGAAGAAATCGCTTTCAAGATGATACGAACCAATGTTTCTCATGTGCTT	3357
QY	3567	GGCAGTTGGATGACATAAAGAAAAAACCCCTAGGAAGTTTGTTCCTCGAATGACAAACATT	3626
DB	3358	GGTCAGTTGGATGACATCAGAAAAAAACCCCAAGGAAGTTGCTTCTCGAATGACAAACATT	3417
QY	3627	GACCACAAATCAATAAGATGCTCAGACAGTGAAGCGTCTCTCAGGAGACTCTATGAATCC	3686
DB	3418	GACCACAAACCAATAAGATGCGCCGACAGTGAAGCGTCTCTCAGGAGACTCTATGAATCC	3477
QY	3687	ATGTTCCCATACCTTCCCAATTTGAACTGCGCAAGAGAGTATCGAAACCGTTTCTCTCAT	3746
DB	3478	ATGTTTCCCATACCTTCCCAAGTTTGAGCTGCCAAGAGAGTATCGAAACCGCTTCTCGAC	3537
QY	3747	ATGCATGAGCTGCAGGAATGAGGGCTTATCGACACAAATTAAGATTTTGGACCCCATGTT	3806
DB	3538	ATGCATGAGCTGCCAAGAAATGCGGGCATATCGAGCAAGCTGAAGTTTGGACCCACTGC	3597
QY	3807	GTACTAGCAACATTGATTATGTTTACTATATTCTCATTTTTTGTCTGAGCAGTTAATTGCA	3866
DB	3598	GTACTAGCAAGTTGATTATATTACTATATTCTCATTTTTTGTCTGAACAGATAATTGCT	3657
QY	3867	CTTAAGCGGAAGATATTTCCCGAAGGAGGATACACAAAGAGCTAGTCCCAATCGAATC	3926
DB	3658	CTGAAGCGAAAGATATTTCCCGAGAGGAGATACACAAAGAGCTAGTCCAGACCCGAATC	3717
QY	3927	AGAGTATAGAAGATCTTCATTTGAAACCACTACTACCTCAGCATTTACTGAGCATTTTAAA	3986
DB	3718	AGGGTGTAGAAGATCTTCATTTGAAAGTCACCTACCTTAGCATCT-GTGAACATCTCCCT	3776
QY	3987	ACTCAGCTTCCAGAGATGCTTTGTGATGTGATGCTTA-GCAGTTTGGCCCCGAGAAGG	4045
DB	3777	CCTCGACACCACACGCGGAGTCCCTGTGATGTGGCACAGAGCAGCCCTCGTGGGAGCAAGG	3836
QY	4046	AAAATATCCAGTACCATGCTGTTTGTGGCATGATATAGCCCACTGACCTAGGAATTATT	4105
DB	3837	GACATCGTGCAGACCGGGTCTCTCTGCAATGGGAAGAGGCCCACTGACCTGGAATTATT	3896
QY	4106	TAACCAACCCACTGAAACHTGTGTGCGAGCAGCTCGAACTGATTTTACTTTTTAAAGA	4165
DB	3897	----CAGCACACTAAGAACCTGTGTCATAGC----TTGTACAGCTTGTACTTTTAAAGG	3948
QY	4166	ATTTGCTCAGGACCTGTCATCTTTTATAAAAAGGCTCACTGACAAGAGACAGCTGTT	4225
DB	3949	ATTTGCCGAGGACCTGTGGCTTGTGTGACAAACCCCTCCCTGACA-----AGCTGCT	4000
QY	4226	AAATTTCCACAGCAATCATTTGCAGACTAACTTTATTAGGAAGACCTATGCCAGCTGGGA	4285
DB	4001	GGTTTCTTCCCCAGTTACTGCAGCT-----GAGAAACCACTCCATCTTGAA	4049
QY	4286	GTGATTTGCTAAGAGCTCCAGTCTTCGATTTCCAAAGCCTTTGTCTAAAGTTTTCGACTT	4345
DB	4050	GCAAGTCGGAGGGGCCAGCTCTTGCATTTCCAAAGCTTTCCAGCATAAATTTCTGGCTT	4109
QY	4346	TT-----TTTTTTTTCATTTCCCATTTT-----AAGTAGTACTAAGTTAACT	4388
DB	4110	GTCTCTCTCTTTGATTCATTTCCCATTTTTTTTTTAAAAACAATAAGTGGCTACTAAGTT	4169
QY	4389	AGTTATTCTCTGCTCTCGATATACGAATTTGGGATGTCTAAACCTATTTTATAGATGTT	4448
DB	4170	AGTCATTTCTCACTTCTCAAAATAACAAATCAGGATGTCAAAAC--ATTTGTATAGATCTT	4227
QY	4449	ATTTTAAATAATGCAGCAATATCACTCTTATTGCAATATCCTAAATATCAGTTTATTATA	4508
DB	4228	ATTTTAAATATATAGAAGATTACTCTT--TTAGCCTATCTAAATTAATTGATTTTTATTATA	4285
QY	4509	ATATTTAAGACTGTAAATGGTCTTTAAACCACT---AACTACTGGAAGAGCTCAATGATTGA	4565

Db	4286	ACAGCTCAAG-----TGGTCTTTGAACCCCTAACAACTACTGAAGAGCTCGA-GATTGA	4336
QY	4566	CATCTGAAATGCTTTTGTAATTATTGACTTCAGCCCTCAAGAAATGCTATGATTTCACGTGC	4625
Db	4337	CGTTGAAGTGCITTTGAGCTTGTTTAACTCATTTCCCAAGAATACTGTGACCTCGTGTC	4396
QY	4626	AGGTCATAATTTCAACAGGCTAGAGT-----TAGTACTACTTACCAGATGTAATTATGT	4678
Db	4397	GGGCGCTGATTTGCGAAGGCTAGTGTCCAGTAGCAGTCTGCTCACCGGATGTAATTATGT	4456
QY	4679	TTTGGAAATGTACATATTTCAAACAGAGTGCCTCATTTTAGAAATCAGTAGTCTCATGG	4738
Db	4457	CGTGGAAATGTACATA--GAGACAAAGTGCCTCACTTCAGAAATGAGTAGTCTCATGG	4514
QY	4739	CATGGCACATTTACAGTG	4756
Db	4515	CACCAGCGAGTGATGGTG	4532
RESULT 13			
ID	ADD27816		
ID	ADD27816	standard; DNA; 5229 BP.	
XX	AC		
XX	AC		
DT	DT		
XX	XX	15-JAN-2004 (first entry)	
DE	DE	GlcNac-phosphotransferase associated DNA #1.	
XX	XX	mouse; protein phosphorylation; soluble GlcNac-phosphotransferase;	
KW	KW	UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.	
XX	XX		
OS	OS	Mus musculus.	
XX	XX		
PN	PN	US2003119088-A1.	
XX	XX		
PD	PD	26-JUN-2003.	
XX	XX		
PF	PF	21-DEC-2001; 2001US-00023888.	
XX	XX		
PR	PR	21-DEC-2001; 2001US-00023888.	
XX	XX	(NOVA-) NOVAZYME PHARM INC.	
PA	PA		
XX	XX	Canfield W, Kudo M;	
PI	PI		
XX	XX	WPI; 2003-801323/75.	
DR	DR		
XX	XX		
PT	PT	Phosphorylating a protein for treating a patient suffering from a	
PT	PT	lysosomal storage disease e.g. Fabry's disease by contacting the protein	
PT	PT	with a soluble GlcNac-phosphotransferase and producing a phosphorylated	
PT	PT	protein.	
XX	XX		
PS	PS	Disclosure; SEQ ID NO 8; 55pp; English.	
XX	XX		
CC	CC	The invention relates to a method of phosphorylating a protein comprising	
CC	CC	contacting the protein with a soluble GlcNac-phosphotransferase (UDP-N-	
CC	CC	acetylglucosamine) and producing a phosphorylated protein. The method is	
CC	CC	useful for treating a patient suffering from a lysosomal storage disease	
CC	CC	e.g. Fabry's disease. The present sequence represents a GlcNac-	
CC	CC	phosphotransferase associated DNA.	
XX	XX		
SQ	SQ	Sequence 5229 BP; 1485 A; 1242 C; 1199 G; 1302 T; 0 U; 1 Other;	
Query Match 50.4%; Score 2820.8; DB 10; Length 5229;			
Best Local Similarity 79.1%; Pred. No. 0;			
Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;			
QY	150	GGCGGTGAAGGGGTGATGCTGTTCAAGCTCTTCGACAGACAAACCTATACCTGCTGTC	209
Db	1	GGCGGTGAAGGGGTGATGCTGTTCAAGCTCTTCGACAGACAGACCTATACCTGCTATCC	60

Db 2221 CTGCTAAGGTCTTCTCTGGGAAATTCTACATAGATATAAAATAAAACCTCAAGCT---AGG 2277
Qy 2427 ACAGATGAACAAATGACAGATTGGTGGCTCCACAGGAAACACAGGTTTCATAAAGCATC 2486
Db 2278 ACCGATGAACAAAGAGCAACTGGAGTCCACAGGAAACCTTCTCACAGACG--- 2333
Qy 2487 TTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGACAGAGTTGACCTTTTCTGCGAGTGAGT 2546
Db 2334 --TCCACATGGCTTTGCTGGTGAACACACAGATCAGAGAGATGACTGCCCCACAGAGACA 2391
Qy 2547 GTAAAGTGAATGTCATGACAGGTCAGAGGTGAGATCCACCCCTGGACTTGGAGACACAGCA 2606
Db 2392 GTGACCGTGAAGGCGGTGACACGCTTTGAATCCACCCCGGTGTTGGAGACCAATGCA 2451
Qy 2607 AGATTAGAGTGAACATCACACCCAAACCAATAGGCGGAATGTGACAAAAGAAAG 2666
Db 2452 AGATT-----GGCCAGCTTACATAGCGGTGACTGTGTCCAAAGAGAC 2496
Qy 2667 CCCCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAGAAAGAAATCACAGG 2726
Db 2497 CTTTCACCGCTGATCGTTCCCCAGAAACCACTTGCACAAAGAGAGAGAGTGCACAGG 2556
Qy 2727 AAAGAAAAGAGACAGTAGAGTGAAGGAAATGCTGAAATCACATAGGCGTTACTGAA 2786
Db 2557 GCAGAA-----GCCAATGCTGTACCTGTAAAGAG 2586
Qy 2787 GTGTTACTTGAAGAAAGCTGCAGCAITPACAGATAGTTACTTTGGGCTTTTTCGCATGG 2846
Db 2587 TTAGTGCCTGGCAGACGGTTGCAGC-----AGAATTATCCAGGCTTTTTCGCCCTGG 2637
Qy 2847 GAGAAAAAAGTATTTCCAGATCTTCTCGACGAGAGAGTCAATTTGAAGACACAATTG 2906
Db 2638 GAGAAAAAAGTATTTCCAGACCTTCTTATGAGGAGAGTCAATTTGAAGACCCAGTTG 2697
Qy 2907 GCATACCTTCACTGATGACAAATACTCGGAGGCAACTAAAGATACATTTGCGAGATTCC 2966
Db 2698 GCGTACTTTAGACCGGCAACATACCGGAGGCACTTAAAGATACATTTGCGAGACTCC 2757
Qy 2967 CTCAGATATGAATAATTTCTAATAAGCAAGTTTGAATTCATACCGGAAAGTCCCT 3026
Db 2758 CTCGATACCTCAATAAATAATTTCTCAACAGCAAGTTTGGATTCACATCCAGGAAAGTCCCT 2817
Qy 3027 GTCACATGCTCAGATGACCGGATTTGATGACAGAACTGCAAGATATGTCCT 3086
Db 2818 GCACATATGCGCCACATGATTGACAGGATCGTTATGCAAGAACTCCAAAGATATGTTCCCT 2877
Qy 3087 GAAGAATTGACAAGACGTCATTTCAAAAGTGGCCATCTGAGGATATGACGTTTGGC 3146
Db 2878 GAAGAATTGACAAGACTTCATTTCAAAAGTGGCTCACTCTGAGGACATGACGTTTGGC 2937
Qy 3147 TTCTCTTATTTTATATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTT 3206
Db 2938 TTCTCTTATTTATTTATCTCATGAGTGCAGTGCAGCCCTCAATATTTCCCAAGTCTTT 2997
Qy 3207 GATGAAGTTGATACAGATCAATCTGGTGTCTGTCTGACAGAGAAATCCGAACACTGGCT 3266
Db 2998 CATGAAGTAGACACAGACCAATCTGTGTCTGTCTGATAGGAAATCCGAACACTGGCC 3057
Qy 3267 ACCAGAAATCAGAACTGCGGTAAAGTTTGCAGGATTTGACAGTCTGGAACACATGCTA 3326
Db 3058 ACGAGAAATCAGAACTTAAAGTTCGAGGATTTGACAGGTTTGGAAACACATGTTA 3117
Qy 3327 ATAAATGCTCAAAATGCTTCCCTGATATCAGCAGCTAAATAATTTCCCAACT 3386
Db 3118 ATAAATGCTCAAAATGCTTCCCTGATATATCTCAACTCAACACATCCACCGACT 3177
Qy 3387 CAGGAATCCTACTATGATCCCAACCTGCCCGGTCACTAAAGTCTAGTAACAACTGT 3446
Db 3178 CAGGAAGCATACTAGACCCCAACCTGCTCCGTCACCTAAGAGTCTTGTCCACCACTGT 3237
Qy 3447 AAACAGTAACTGACAAATCCACAAAGCATATAGGACAAACAAATATAGTTTGA 3506
Db 3238 AAGCCAGTAAGTGAAGATCCACAAAGCCTATAAAGACAAAGAAATACAGGTTTGA 3297

Qy 3507 ATCATCGGAGAGAAATCGCTTTTAAATGATTCTGTACCAACGTTTCTCATGTGGTT 3566
Db 3298 ATCATCGGAGAGAGAAATCGCTTTCAAGATGATACGAAACCAATGTTTCTCATGTGGTT 3357
Qy 3567 GGCCAGTTGATGACATAGAAACCACTAGAGAGTTTGTTCCTGTAATGACAACTT 3626
Db 3358 GGTCACTGGATGACATCAGAAAAACCCAGAGAGTTTGTTCCTGTAATGACAACTT 3417
Qy 3627 GACCAATCATAAAGATCTCAGACAGTGAAGGCTTCTCAGGACACTTCTATGAATCC 3686
Db 3418 GACCAATCATAAAGATGCCGACAGTGAAGGCTTCTCAGGACACTTCTATGATCC 3477
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAACTGCCAAAGAGAGTATCGAAACCGTTTCTTCTAT 3746
Db 3478 ATGTTCCCATACCTTCCCAATTTGAGCTGCCAAAGAGAGTATCGAAACCGTTTCTTCTGAC 3537
Qy 3747 ATGATGAGCTCGAGAAATGGAGGCTTATCGAGACAAATGAAAGTTTGGACCCATCT 3806
Db 3538 ATGATGAGCTCGAGAAATGGAGGCTTATCGAGACAAATGAAAGTTTGGACCCATCT 3597
Qy 3807 GTACTAGCAACATGATGATTTTACTATATTTCTATATTTCTCATTTTGTCTGAGAGTAAATGCA 3866
Db 3598 GTACTAGCAACATGATGATTTTACTATATTTCTATATTTCTCATTTTGTCTGAAACAGATAATGCT 3657
Qy 3867 CTTAAGCGGAGATATTTCCAGAGAGAGATACACAAAGAGAGTATCTCAGCACTTAAATGCA 3926
Db 3658 CTTAAGCGGAGATATTTCCAGAGAGAGATACACAAAGAGAGTATCTCAGCACTTAAATGCA 3717
Qy 3927 AGAGTATGAGATCTTCAATTTGAAACCATCTACCTCAGCACTTAAATGCACTTAAATGCA 3986
Db 3718 AGGATGAGAGATCTTCAATTTGAAAGTCACTTACCTTAGCATCT--GTGAAACATCTCCCT 3776
Qy 3987 ACTCAGCTTTCACAGAGATGCTTTTGTGATGTGATGCTTA--GCAGTTTGGCCGAGAGAG 4045
Db 3777 CCTCGACACACAGCGAGTCCCTGTGATGTGCAAGAGGAGAGGCTCTGGGAGAGAG 3836
Qy 4046 AATATATCAGTACATGCTGTTTGTGCAATGATATAGCCCACTGACTAGCAATAT 4105
Db 3837 GACATCGTGCAGACCGGTTCTTCTCAATGGAAGAGAGAGCCACTGACTGGAATAT 3896
Qy 4106 TAACCAACCACTGAAACCTTGTGTCTGAGAGAGTCTGAACTGATTTTACTTTTAAAGA 4165
Db 3897 ----CAGCACATAAGAACCTGTGCAATAGC----TTGTACAGCTTGTACTTTAAAG 3948
Qy 4166 ATTTGCTATGACCTGCTCATCTTTTATAAAGGCTCATGACAGAGAGAGAGTGT 4225
Db 3949 ATTTGCGAGAGACCTGTGGGCTTGTGACAAACCCCTCCCTGACA-----AGCTGCT 4000
Qy 4226 AATTTCCACAGCAATCATTTGAGAGTAACTTTATAGGAGAGGCTATGCCAGCTGGGA 4285
Db 4001 GGTTCCTTCCCGAGTACTGAGACT-----GAGAAACAGTCCATCTTGAA 4049
Qy 4286 GTGATGCTAAGAGGCTCCAGTCTTTGCAATCCAAAGCCTTTTGTGTAAGTTTGCAT 4345
Db 4050 GCAAGTGGGAGGCGCCAGTCTTTGCAATCCAAAGCTTTCCAGCATATTTCTGGCT 4109
Qy 4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTACTAGTTAACT 4388
Db 4110 GTCTCTCTCTTGTATCCATTTTCCCAATTTT-----TAAAGAAACAAATAGTGGCTACTAGTT 4169
Qy 4389 AGTATTTCTTGTCTGAGTAAACGAATGGGATGCTTAAACCTATTTTATAGATGTT 4448
Db 4170 AGTCAATCTCACITCTCAAAATAACAAATCAGGATGTCAAAGC--ATTTGATAGATCT 4227
Qy 4449 ATTTAAATATGAGCAATATCACCTCTTATGCAATACCTTAAATATAGTATTTATTA 4508
Db 4228 ATTTAAATATATAGAACGATTTACTTCT--TTAGCCTATCTAAATATTTGATTTTATTA 4285
Qy 4509 ATATTTAAGACTGTAATGCTTAAACCACT---AACTACTGAGAGGCTCAATGATTA 4565
Db 4286 ACAGTCAAG-----TGTCTTGAACCGCTAACACTACTAGAGAGTCCA--GATTGA 4336

QY	3447	AAA	CCAGTAACTGACAA	AAATCCACAAGCATAT	TAAGGACAAA	AAACAAATATAGT	TTTGAA	3506
DB	3238	AAGCCAGTAACTGACA	GATCCACAAGCCCTAT	TAAGACAAGAA	CAAAATACAGG	TTTGTAA	3297	
QY	3507	ATCATGGGAGAA	GAAATCGCTTTTAA	AAATGATTCGTACCA	ACGTTTCTCAT	GTGCGTT	3566	
DB	3298	ATCATGGGAGAA	GAAATCGCTTTCA	AGATGATA	CGAACCATGTTTCTCAT	GTGCGTT	3357	
QY	3567	GGCCAGTTGGAT	GATACATAAGAAAA	AAACCCCTAGGA	AGTTTGTTCCTGA	ATGACAA	CAAT	3626
DB	3358	GGTCAGTTGGAT	GATACATCAGAAAA	AAACCCCAGGA	AGTTTGTTCCTGA	ATGACAA	CAAT	3417
QY	3627	GACCACAATCAT	AAAGATGCTCAGACAGTGA	AGGCTGTTCTCAGG	GA	CTTCTATGA	TCC	3686
DB	3418	GACCACAACCA	TAAGATGCCCCGACAGTGA	AGGCTGTTCTCAGG	GA	CTTCTATGA	TCC	3477
QY	3687	ATGTTTCCCCAT	CACTTCCCAATTTGA	ACTGCCAAGAGAGTATCGAA	ACCGTTTCCCTTCAT			3746
DB	3478	ATGTTTCCCCAT	CACTTCCCAATTTGAGCTGCCA	AGAGAGTATCGAA	ACCGTTTCTG	CA		3537
QY	3747	ATGCATGAGCTG	CAGGAATGAGGGCTTATCGACACAA	ATGAAGTTTTCGAC	CCATGTT			3806
DB	3538	ATGCATGAGCTC	CAAGAAATGGCGGCATATCGAGACA	AGTGAAGTTTTCGAC	CCCACTGC			3597
QY	3807	GTACTAGCAACAT	TGATTGTTTACTATATTTCTCAT	TTTTTGTGAGCAGT	TAATGCA			3866
DB	3598	GTACTAGCAAGT	TGATTATTTACTATATTTCTCAT	TTTTTGTGAGCAGAT	TAATGCT			3657
QY	3867	CTTAAAGCGGA	GATATTTCCCAAGAGAGAGATACACA	AAAGCTAGTCCCAATCGA	ATC			3926
DB	3658	CTGAAGCGGA	GATATTTCCCAAGAGAGAGATACACA	AAAGCTAGTCCCAAGC	CAATC			3717
QY	3927	AGAGTATAGAA	GTCTTCATTTGAAAA	ACCATCTACCTCAGCATTTACTGAGC	ATTTTAA			3986
DB	3718	AGGTTGTAAGA	ATCTTCATTTGAAAGTCACTTACCTTAGCATCT					3776
QY	3987	ACTCAGCTTCA	CAGAGATGCTTTTGTGATGTGATGCTTA					4045
DB	3777	CCTCGACACAC	AGCGGAGTCCCTGTGATGTGGCACAGAG	CGACCTCGTGGG	GAGAAG			3836
QY	4046	AAAATATCCAGT	TACCATCTGTTTGTGGCATGATATAGCCCACTGAC	TAGAGAA	TTATTT			4105
DB	3837	GACATCGTCAG	ACCGGTTCTCTCGAAATGGGAAGAGCCCACTGAC	CTGAGAA	TTATTT			3896
QY	4106	TAAACCAACCC	ACTGAAAACTTGTGTCGAGCAGCTCTGA	ACTGATTTTACTTTTTAA	AGA			4165
DB	3897	-----	CAGCACATPAAGAACCTGTGTCAATAGC	----				3948
QY	4166	ATTGTCTCAT	TGGACCTGTATCTCTTTTATAAAAGGCTCACTGA	CAGAGACAGCTGTT				4225
DB	3949	ATTTGGCGA	AGGACCTGTGCGCTGTTTGACAAA	CCCTCCCTGACA	-----			4000
QY	4226	AAITTTCCACAG	CAATCATTTGCAGACTTA	CTTTTATTAGGAGAA	CGCTATGCCAGCTGGGA			4285
DB	4001	GGTTCTCTCC	CCAGTTACTGCAGACT	-----				4049
QY	4286	GTGATTTCT	TAAGGCTTCCAGTCTTTTGCAATCAAA	CGCTTTTGTATAAGTTT	TTCGACTT			4345
DB	4050	GCAAGTGC	GGAGGGGCCCACTCTTTGCAATTCAAA	AGCTTTCCAGCATATTTCTG	CGCTT			4109
QY	4346	TT-----	TTTTTTTTCATTTCCCATTTTTT	-----				4388
DB	4110	GTCTCTCCT	TTTGTATTCATTTTCCATTTTTTTTTTAAAA	AAACAATAGTGC	TACTAAGTT			4169
QY	4389	AGTATTTCT	TGCTCTGAGTATAACGAATTTGGATGTCTAA	ACCTATTTTTTATAGATGTT				4448
DB	4170	AGTCATTTCT	CACTTCTCAAAAATAACAAATCAGGATGTCA	AAAC	-			4227
QY	4449	ATTTAAATAT	TGAGCAATATACCTCTTATTCAGCAATAC	CTTAATATAGCTTTTATTA				4508
DB	4228	ATTTAAATA	TATAGAACGATTA	CTTCT--TTAGCCTATCTAA	ATTTATTTATTTA			4285
QY	4509	ATATTTTAAG	ACTGTAAATGTCTTAAACCACT	---				4565

Db	4286	ACAGCTAAG-----TGGTCTTTGAACCGCTAACCACTACTGAAGAGCTCGA-GATTGA	4336
QY	4566	CATCTGAAATGCTTTGTAATTATTGACTTCAGGCCCTTAAGAATGCTATGATTTCACGTGC	4625
Ds	4337	CGTTGAAGTGCCTTGAGCTTGTTAACTCATTTCCCAAGAATACTGTGACCTCGTGTGC	4396
QY	4626	AGGTCTAATTTCAACAGGCTAGAGT-----TAGTACTACTTACCAGATGTAATTATGT	4678
Ds	4397	GGGCTGATTCGGAAGGGCTAGTGTCAAGTAGCAGTGTCTCTCACCGGATGTAATTATGT	4456
QY	4679	TTTGGAAATGTACATATTTCAAACAGAGTGCCTCATTTTAGAAATGAGTAGTCTGATGG	4738
Ds	4457	CGTGGAAATGTACATA--CAGACAAAGTGCTTCACTTCAGAAATGAGTAGTCTGATGG	4514
QY	4739	CACTGGCACATTACAGTG	4756
Ds	4515	CACCAGCGAGTGATGGTG	4532
RESULT 15			
AAD62652			
ID	AAD62652 standard; DNA; 5229 BP.		
XX			
AC	AAD62652;		
XX			
DT	15-JAN-2004 (first entry)		
XX			
DE	Mouse DNA #1 used to illustrate the method of the invention.		
XX			
KW	Glycoprotein; lectin; lysosomal storage disease; Gastrointestinal;		
KW	N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse; ds.		
XX			
OS	Mus musculus.		
XX			
PN	US2003124653-A1.		
XX			
PD	03-JUL-2003.		
XX			
PF	21-DEC-2001; 2001US-00023890.		
XX			
PR	21-DEC-2001; 2001US-00023890.		
XX			
PA	(NOVA-) NOVAZYME PHARM INC.		
XX			
PI	Canfield WM;		
XX			
DR	WPI; 2003-810985/76.		
XX			
PT	Producing a glycoprotein with reduced complex carbohydrates by culturing		
PT	the lectin resistant mammalian cell expressing the glycoprotein for		
PT	treating lysosomal storage disease.		
XX			
PS	Disclosure; Page 20-23; 46pp; English.		
XX			
CC	The present invention provides a method of producing a glycoprotein		
CC	having reduced complex carbohydrates by culturing the lectin resistant		
CC	mammalian cell expressing the glycoprotein. The method is useful for		
CC	producing a glycoprotein with reduced complex carbohydrates for treating		
CC	lysosomal storage disease. The present invention is also useful in gene		
CC	therapy. The present sequence is mouse DNA used to illustrate the method		
XX	of the invention		
SQ	Sequence 5229 BP; 1485 A; 1242 C; 1199 G; 1302 T; 0 U; 1 Other;		

QY 210 CACAGTATGGCTCTAGCTGTCTTCTGGCGCTGTTGTACCATCGTCTCGCCTTC 269
Db |||||
QY 61 CACAGTATGGCTCTAGCTGTCTTCTGGCGCTGTTGTACCATCGTCTCGCCTTC 120
Db |||||
QY 270 CAGTTGAGAGAGTGGTCTGGAATGGAGCCGAGATCAATACCATGTTTGTGATTC 329
Db |||||
QY 121 CAGTTGAGAGAGTGGTCTGGAATGGAGCCGAGATCAGTACCATGTTTGTGATTC 180
Db |||||
QY 330 TATAGAGACAATATGCTGGAAAGTCTTTTTCAGAAATGGCTTGTGTCCTCCATCCGAT 389
Db |||||
QY 181 TACAGAGACAATGCTGGGAATCTTTTTCAGAAATCGGCTCTGTCTGCCCATGCCAATC 240
Db |||||
QY 390 GACGTTGTTTACCTGGTGAATGGCACAGATCTTGAATCTTGAAGAACTACAGAG 449
Db |||||
QY 241 GACGTTGTTTACCTGGTGAATGGCACATGACCTTGAATCTTGAAGAGCTACAGAG 300
Db |||||
QY 450 GTCAGAGACAGATGGAGAGGAGCAGAAAGCAATGAGAAATCTTGGGAAACACA 509
Db |||||
QY 301 GTCCGAGACATGGAGAGAGCAGAGAGCCATGGGGAACCCCTGGGAAGACACA 360
Db |||||
QY 510 ACGGAACCTACTAAGAGAGTGAAGAGCAGATGTAGAGTGTGTGTAACACATCGATTAAG 569
Db |||||
QY 361 ACCGAACCGACAAAGAGTGAAGAGCAGCTGGAATGTCTGTGACGCACTGCATTAG 420
Db |||||
QY 570 GTGCCAATGCTGTACTGGAACCGCCTGCCAGCAACATCACCCCTGAAGAGCTGCCA 629
Db |||||
QY 421 GTGCCAATGCTGTACTGGAACCGCCTGCCAGCAACATCACCCCTGAAGAGCTGCCA 480
Db |||||
QY 630 TCTCTTTTATCTCTTTTCTGCTGCAATGATGATGATGATGATGATGATGATGATGAT 689
Db |||||
QY 481 ACCCTTTACCCATCTTTTCCAGCTGCAAGCAGATGTTTCAATGTTGCAACCAAAAT 540
Db |||||
QY 690 CTTCTACCAATGCTCTAGTGTGTTTGTGCACTACTAAGAGTGTGAAAGTGTGCCAC 749
Db |||||
QY 541 CGTCTACCAATGCTCTAGTGTGTTTGTGCACTACTAAGAGTGTGAAAGTGTGCCAC 600
Db |||||
QY 750 TCTGACCTGCTTAAAGGAAATAGCAGACAGATGATGAGGGGTACTTGCACACAGAT 809
Db |||||
QY 601 GCTGACCTGTTTAAAGGAGCCAGCAACAGATGTTTGAAGAGCTTCTTGCACACAG 660
Db |||||
QY 810 AAAGAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 869
Db |||||
QY 661 AAAGAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db |||||
QY 870 ACATTTCAAGGAAACAAATCACTAAACAAATTTGCCAGAAATCTTCTCTCTCTCTCT 926
Db |||||
QY 721 ACCTTTCAAGGAGCAGTCACTGAAGCAAAAGCTGCAAGAAAGCTTCTCTCTCTCTCT 780
Db |||||
QY 927 GTCAAACTGCTGAGTGTATTCAGAGGCGAGTGTAGCGCTTCTTAAACTGAATAACCCC 986
Db |||||
QY 781 ATAAAGCTGTTGCGCTGTACTCGAGGCGCAGTGTCTCTTCTGAAATGGAATATCCC 840
Db |||||
QY 987 AAGGATTTTCAAGAAATGGAATGAACAACTAAGAGAAACATGACCATGATGGGAGGNA 1046
Db |||||
QY 841 AAGGATTTTCAAGAGCTGAACAAAGCAGACCAAGAGAAACATGACCATGATGGGAGGNA 900
Db |||||
QY 1047 CTGACCAATAGTCTGATATTTATATGGGATCTGAGGCGCATCAGCAGCTCTAAGCAG 1106
Db |||||
QY 901 CTGACCAATAGGCTCTGCTGTGAGGACCTTGTGAGGCGCATCAGCAGCTCTAAGCAG 960
Db |||||
QY 1107 GATGAAGACATCTCTGCCAGTCTGTTTGAAGATAACGAAGAACTGAGTCTCTAATGCGA 1166
Db |||||
QY 961 GATGAAGACATCTCTGCCAGTCTGTTTGAAGATAATGAAGAGCTGAGTCTCTGCTGGA 1020
Db |||||
QY 1167 TCTATCGAGAGGATGACCATGAGTGTGGAATATTTTCAATGTCACCAACGGGAGAT 1226
Db |||||
QY 1021 TCTATCGAGAGACAGCGCATGGGTACGGAATATTTTCAATGTCACCAACGGGAGAT 1080
Db |||||
QY 1227 CCATCTGCTGCACTCTGACAACTCTGAGTGAACATAGTAACACACAGGATGTTTT 1286
Db |||||
QY 1081 CCATCTGCTGCACTCTGACAACTCTGAGTGAACATAGTAACACACAGGATGTTTT 1140
Db |||||
QY 1287 CGAAATTTGAGCAGCTGTGCTACCTTGTAGTTCACCTGTATTTGAAAGTCAATTCATCGC 1346
Db |||||

Db |||||
QY 1141 CAAAATCTGAGCAGCTTGCCTACTTTCAGTTCCTCTGCTATTGAAAGTCACATTCACCGC 1200
Db |||||
QY 1347 ATCGAAGGCTCTCCAGAGGTTTATTTTACCTAAATGATGATGATGATGATGATGATGAT 1406
Db |||||
QY 1201 ATCGAAGGCTCTCCAGAGGTTTATTTATCTAAATGACATGATGATGATGATGATGATGAT 1260
Db |||||
QY 1407 GTCTGCGCAGATGATTTTACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCT 1466
Db |||||
QY 1261 GTCTGCGCAGATGATTTTACAGCCTCCAAAGGTCAAAAGGTTTATTTGACATGGCCT 1320
Db |||||
QY 1467 GTCCAAAATGTCGCCAGAGGCTGCCAGGTTCTCTGATTTAAGATGGCTATTGTGACAG 1526
Db |||||
QY 1321 GTCCAAAATGTCGCCAGAGGCTGCCAGGCTCTCTGATTAAGAGAGCGCTATTGTGATAAG 1380
Db |||||
QY 1527 GCTGTAAATTAATTCAGCTGCGATTTGGGATTTGGGATTTGCTCTGAAAACAGTGGAGG 1586
Db |||||
QY 1381 GCTGTAAATTAATTCAGCTGCGATTTGGGATTTGGGATTTGCTCTGAAAACAGTGGAGG 1440
Db |||||
QY 1587 AGTGGCTTATTCAGAGAGTGGAGTACTTGGAGTATTGGAGTTGGACACCCCTGGCAG 1646
Db |||||
QY 1441 AACCGGTTTGTTCGAAGAGTGGGCTTACCGGAAATTTGGAGCTGGACAGCAGCTGGCAG 1500
Db |||||
QY 1647 TTTGGTGGAGGATTAACAGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTC 1706
Db |||||
QY 1501 TTTGGTGGAGGATTAACAGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTC 1560
Db |||||
QY 1707 GCTGATAAGTCTGTGACCAAGCATCAATGCTCTGCTGTGGGTTTGTGATGTCGCGAC 1766
Db |||||
QY 1561 GCTGACAAATTTCTGTGACCAAGCTTAACTGCTTATCTGCGGTTTGTGATGTCGCGAC 1620
Db |||||
QY 1767 TGTGGGCAAGATCATTTTTCATGATTTGATTAAGTGTCTCTTCTCCCAACAGCAGCTCAC 1826
Db |||||
QY 1621 TGTGGCAAGATCATTTTTCATGATTTGATTAAGTGTCTCTTCTCCCAACAGCAGCTCAC 1680
Db |||||
QY 1827 TATATTTATTTCCAAAAGTGAATGCTGCTCTTATTTTTCAGCTTTTGCAGAACTAGCAGAAAGA 1886
Db |||||
QY 1681 TATGTTGCTCCCAAGGTTGAATACCTGCTTATTTTTCAGCTTTTGCAGAACTAGCAGAAAGA 1740
Db |||||
QY 1887 GAGTTTGAAGGTCCTTATGATGATCAATTCATTAATTCGACATGCTTCTTATTCGACAAAG 1946
Db |||||
QY 1741 AGAATTTGAAGGACCTTACAGCAGCAACCCCTCATCCGCGCACGCTTCCATTTGCAAAACAAG 1800
Db |||||
QY 1947 TGAAACCACTCCACCTCATTAATGACAGTGGAAATGAATGCGCAACCAATACATTTTAAAT 2006
Db |||||
QY 1801 TGAAGAACCTTACCTGATTAATGCGGGGGGATGAAAGCCACCAAGATCTATTTTAAAC 1860
Db |||||
QY 2007 CTCACTTTTCAAAATACAAACGATGAAGATTCAAAATGCAATACAGATAACAGTGGAGTGGAC 2066
Db |||||
QY 1861 CTCACTTTTCAAAACGCGACGACGAGATTCAGATCCAGATCCAGATAGAGTGGAG 1920
Db |||||
QY 2067 ACAGGAGGAGCAAACTGAAATTTCTCGGCGGAGAGGTTTACGAAATTTAGTTAGT 2126
Db |||||
QY 1921 ACAGGAGGAGGCGGCAAACTGAAATTTCTAACCCAGAGGCTTATGAAAGTTTGGTTAGC 1980
Db |||||
QY 2127 CCATTAACCTTCTTCCAGAGGCGGAAATCTCTTTTGGAGATATTCGAAAGGAAACAGC 2186
Db |||||
QY 1981 CCAGTGACACTCTTCTCTCAGGCTGAGCTCTCTTTTGAAGATGTCCTCCCAAGAGAAACAGC 2040
Db |||||
QY 2187 TTCCGAGGTTTGAAGACATGATTTTAACTCAACAGGAGAGCCAGGAGAGGAGTGAAG 2246
Db |||||
QY 2041 TTCCCAAGATCAGGAGACATGATGTAATTTGCAACAGGAGGATTTCCAAAGAGGAGTGAAG 2100
Db |||||
QY 2247 ATTCCCTGTGTAATTTTTCACCTCTTCCAAAAGAGCGCCAGTTGAGTCTCAATACCTTG 2306
Db |||||
QY 2101 ATCCCGGGTAAATTTTTCACCTCTTCCCAAGAGGCGGAGGCTGAGGCTGAGCACTTG 2160
Db |||||
QY 2307 GATTTGCAACTGGAACATGAGACATCTCTTTGAAAGGATTAACATTTTGTCCAGTCAAGC 2366
Db |||||
QY 2161 GATTTGCAACTGGAACATGAGACATCTCTTTGAAAGGATTAACATTTTGTCCAGTCAAGC 2220
Db |||||
QY 2367 TTGCTGATGATCTTCTGATGAACTCAGCAGTGTGTAATTTTAAAGTCAAGCTAATA 2426
Db |||||

Db 2221 CTGCTAAGGCTCTTTCTCGGGAAATTCAGTAGATACTAAATAAACCCTCAAGCT---AGG 2277
Qy 2427 ACAGATGAACAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATC 2486
Db 2278 ACCGATGAACAAAGGCAACCTGGAGTCCACAGGAARACCTTCTCAGAGC--- 2333
Qy 2487 TTGCCAAACAGCTTGAAGAGTGTGAAAGATTGACAGGTTGACTTTTCTCGAGTGAGT 2546
Db 2334 --TCACATCGCTTTGCTGGTGAAACACAGATCAGAGATGAGTGTGCCCCAGCAGAGACA 2391
Qy 2547 GTAAAGTGAATGGTCATGACCGAGGTGAGATCCACCCCTGGACTTGGAGACCACAGCA 2606
Db 2392 GTGACCGTGAAGGCGGTGACACGCTTTGAAATCCACCCCGGTGTGGAGCCAAATGCA 2451
Qy 2607 AGATTTAGAGTGAATACTCACACCAAAAAACCATAGCGGAAATGTGACAAAAAGAAAAG 2666
Db 2452 AGATT-----GGCCAGCCTACACTAGGCGTGAAGTGTGTCAAAAGAGAAC 2496
Qy 2667 CCCCATCTCTGATTGTTCTCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGG 2726
Db 2497 CTTTCACCGCTGATGTTCCCCCAAGAACCACTTGCCAAAGAAAGAGAGAGTGACAGG 2556
Qy 2727 AAAGAAAAAGAGAACAGTAGAATGGAGAAAAATGCTGAAATACATAGGCGTTACTGAA 2786
Db 2557 GCAGAA-----GGCAATGCTGTACCTGTAAAGGAG 2586
Qy 2787 GTGTTACTTGAAGAAAGCTGAGCATTTACAGATAGTTACTTGGGCTTTTGGCCATGG 2846
Db 2587 TTAGTGCCTGCGACGCTTGCAGC-----AGAAATTACAGGCTTTTGGCCCTGG 2637
Qy 2847 GAGAAAAAAAGTATTTCCAGATCTCTCGACGAAGAGATCATTTGAGAGACAAATG 2906
Db 2638 GAGAAAAAAAGTATTTCCAGACCTTCTGATGAGGAGAGTCTTTGAGACCAGTTG 2697
Qy 2907 GCATCTTCACTGATGAGAAAAATACCTGGAGGCAACTTAAAGATACATTTGCAAGATCC 2966
Db 2698 GCGTACTTTACAGACCGCAACATACCGGAGGCAACTTAAAGATACATTTGCAGACTCC 2757
Qy 2967 CTCAGATGATAAATAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGAAAGTCCCT 3026
Db 2758 CTCGATACGTCATAAATAATTTCTCAACAGCAAGTTTGGATTCACATCAGAAAGTCCCT 2817
Qy 3027 GCTCAGATGCTCAGATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCT 3086
Db 2818 GCACACATGCGGCACATGATTGACAGGATCGTTATGCAAGAACTCCAAAGATATGTTCCCT 2877
Qy 3087 GAAGNAATTTGACAGACGTCATTTACAAAGTGCCTCTGAGGATATGCAAGTTGCC 3146
Db 2878 GAAGAAATTTGACAGACGTCATTTACAAAGTGCCTCTGAGGACATGCAAGTTGCC 2937
Qy 3147 TTCTCTTATTTTATTTATCTCATGAGTGCAGTGCAGGCACTGAATATATCTCAAGTCTTT 3206
Db 2938 TTCTCTTATTTTATTTATCTCATGAGTGCAGTGCAGGCACTGAATATATTTCCAGTCTTT 2997
Qy 3207 GATGAAGTTGATACAGATCAATCTGGTGTCTGTCACAGAGAAATCCGAAACATGGCT 3266
Db 2998 CATGAAGTAGACACAGACCAATCTGGTGTCTGTCATAGGGAATCCGAAACATGCTGGCC 3057
Qy 3267 ACCGAAATTCAGAACTGCGGTTAAGTTTGGAGATTTGACAGTCTGGAACACATGCTA 3326
Db 3058 ACGGAATTCAGACCTTACCTTTAAGCTTGCAGGATTTGACAGTTTGGAAACATGTTA 3117
Qy 3327 ATAAATTTGCTCAAAAAATGCTCCCGCTAATATCACTCAACTCAACACATCCCAAGGACT 3177
Db 3387 CAGGAATCTACTATGATCCCAAGCTGCGGCTCACTAAAGTCTAGTAACAAACTGT 3446
Qy 3178 CAGGAATCTACTAGACCCCAAGCTGCGGCTCACTAAGAGTCTTTGTCACAACTGT 3237
Qy 3447 AAACCGATTAAGTCAAAAAATCCAAAGCATATTAAGGACAAAAACAAATATAGGTTTGA 3506
Db 3238 AAGCCAGTAACTGACAGATCCAAAGGCTTATAAGACAAAGAAATACAGGTTTGA 3297

Qy 3507 ATCATGGGAGAGAGAAATCGCTTTTAAATGATTTCGTACCACGTTTCTCATGTGGTT 3566
Db 3298 ATCATGGGAGAGAGAAATCGCTTTCAAGATGATACGAACCAATGTTTCTCATGTGGTT 3357
Qy 3567 GGCAGTTCGATGACATTAAGAAAAACCCCTAGAGAGTTTGTGCTGAAAGACAAATTT 3626
Db 3358 GGTCAAGTTGGATGACATCAGAAAAAACCCCAAGAAAGTTGCTTGTCTGAATGACAAAT 3417
Qy 3627 GACCAAAATCAATAAGATGCTCAGACAGTGAAGGTGTTCTCAGGAGTCTTCTATGAATCC 3686
Db 3418 GACCAAAACATTAAGATGCTCCGACAGTGAAGGTGTTCTCAGGAGTCTTCTATGATCC 3477
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAATGCTGCAAGAGAGTATCGAAAAACGTTTCTTCTCAT 3746
Db 3478 ATGTTTCCCATACCTTCCCAATTTGAATGCTGCAAGAGAGTATCGAAAAACGTTTCTTCTGCAC 3537
Qy 3747 ATGCAATGCTCGAGAAATGAGGCTTATCGAGACAAATGAAGTTTGGACCCCATGTT 3806
Db 3538 ATGCAATGCTCGAGAAATGAGGCTTATCGAGACAAATGAAGTTTGGACCCCATGTT 3597
Qy 3807 GTACTAGCAACATTTGATTATGTTTACTATATTTCTCATTTTGTCTGAGCAGTTAATGCA 3866
Db 3598 GTACTAGCAACATTTGATTATATTTACTATATTTCTCATTTTGTCTGAGCAATATGCT 3657
Qy 3867 CTTAAGCGGAGAAATTTCCAGAGAGGAGTATACAAAGAGCTAGTCCCAATCGAATC 3926
Db 3658 CTGAAGCGAAAGATATTTCCAGAGGAGGATACAAAGAGCTAGTCCAGACCGAATC 3717
Qy 3927 AGAGTATAGAAATCTTCTCATTTTGAACCACTACTCTCAGCAATTTACTGAGCAATTTTAAA 3986
Db 3718 AGGTGTAGAGATCTTCTCATTTGAAGTCACTACTCTTAGCATCT-GTGAACATCTCCCT 3776
Qy 3987 ACTAGCTTCAAGAGATGCTTTGTGATGATGATGCTTA-GCAGTTTGGCCCGGAAGAG 4045
Db 3777 CCTCGACACCAAGAGGAGTCCCTGTGATGTGGCAAGAGGAGGCTCGTGGGAGAGG 3836
Qy 4046 AAAATATCAGTACCATGCTGTTTGTGGCATGAATATAGCCCACTGACTAGGAATATT 4105
Db 3837 GACATCGTGCAGACCGGGTTCTTCTGCAATGGAGAGAGGCCACTGACCTTGAATATT 3896
Qy 4106 TAAACAAACCACTGAAAACTTTGTGTGTGAGCAGCTCTGAACTGATTTTACTTTTAAAGA 4165
Db 3897 ----CAGCACACTAAGAACCTGTGTCAATAGC----TTGTACAGCTTGTACTTTTAAAG 3948
Qy 4166 ATTTGCTCATGAGACCTGTCATCTTTTATAAAGGCTCACTGACAGAGACAGCTGTT 4225
Db 3949 ATTTGCGAGAGACCTGCTGGTGTGACAAACCCCTCCCTGACA-----AGCTCT 4000
Qy 4226 AATTTCCCAAGCAATCAATGCACTAACTTTATTAGGAGAGGCTTATGCGAGCTGGGA 4285
Db 4001 GGTTTCTTCCCGAGTACTGAGACT-----GAGAAACCACTGCTCATCTTGAAA 4049
Qy 4286 GTGATTTGCAAGAGCTCGAGTCTTGTGATTCAAAGGCTTTTGTGTAAGTTTGTGACTT 4345
Db 4050 GCAAGTGGGAGGCGGCGGCTTGTGATTCGAATTCGAAGCTTTCCAGCATAAATTTCTGGCTT 4109
Qy 4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTTACTAAGTTAACT 4388
Db 4110 GTCTCTCTTGTGATTCATTTCCCATTTT-----TAAACAACTAAGTGGCTACTAAGTT 4169
Qy 4389 AGTTATCTTGTCTGAGTATAGAAATGCGAGTCTTAAACCTATTTTATATGATGTT 4448
Db 4170 AGTCAATCTCACTTCTCAAAATAACAAATCAGAGTGTCAAAAC--ATTGTATAGATCTT 4227
Qy 4449 ATTTAAATAATGACCAATATCACTCTTATTGCAATACCTTAAATATATGATGTTTATTA 4508
Db 4228 ATTTAAATAATATAGAACGATTAATCTTCT--TTAGCTATCTAAATATATGATTTTATTA 4285
Qy 4509 ATATTAAAGCTGTAATGTTTAAACCACT---AACTACTGAGAGCTCAATGATTGA 4565
Db 4286 ACAGTCAAG-----TGGTCTTGAACCGCTTAACTACTGAGAGCTCGA-GATTGA 4336

QY	4566	CATCTGAATGCTTTTGTAAATTATTGACTTCAGCCCTAAGAATGCTATGATTTCACGTGC	4625
Db	4337	CGTTGAAGTGTCTTTGAGCTTCTTTAACTCATTCCTCCCAAGATACTGTGACCTCGTGTGC	4396
QY	4626	AGGTCTAATTTCACACAGGCTAGAGT-----TAGTACTTACCAGATGTAATTATGT	4678
Db	4397	GGGCTGTATTGCGAAGGGCTAGTGTACGTAGCAGTGTCTCACCCGATGTAATTATGT	4456
QY	4679	TTTGGAATGTACATATTCAACAGAGTGCCTCATTTTAGAAATGAGTAGTGTGATGG	4738
Db	4457	CGTGAATGTACATA--CAGCAAAAGTGCCTCACTTCAGAAATGAGTAGTGTGATGG	4514
QY	4739	CACTGGCACATTACAGTG	4756
Db	4515	CACCAGCGAGTGATGGTG	4532

Search completed: November 21, 2004, 23:56:11
Job time : 2432 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 23:00:01 ; Search time 412 Seconds

(without alignments)

9656.032 Million cell updates/sec

Title: US-10-023-888-3

Perfect score: 5597

Sequence: 1 cggagccgagcgggctccg.....aaaaagtaattttgaaa 5597

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5597	100.0	5597	4	US-09-635-872A-4
2	5597	100.0	5597	4	US-09-636-077A-4
3	5597	100.0	5597	4	US-09-636-060C-4
4	5597	100.0	5597	4	US-09-986-552-4
5	5597	100.0	5597	4	US-09-636-596C-4
6	3761	67.2	3783	4	US-09-638-872A-20
7	3761	67.2	3783	4	US-09-636-077A-20
8	3761	67.2	3783	4	US-09-636-060C-20
9	3761	67.2	3783	4	US-09-986-552-20
10	3761	67.2	3783	4	US-09-636-596C-20
11	3485.4	62.3	3621	4	US-09-638-872A-21
12	3485.4	62.3	3621	4	US-09-638-077A-21
13	3485.4	62.3	3621	4	US-09-636-060C-21
14	3485.4	62.3	3621	4	US-09-986-552-21
15	3485.4	62.3	3621	4	US-09-636-596C-21
16	2820.8	50.4	5229	4	US-09-635-872A-16
17	2820.8	50.4	5229	4	US-09-636-077A-16
18	2820.8	50.4	5229	4	US-09-636-060C-16
19	2820.8	50.4	5229	4	US-09-986-552-16
20	2820.8	50.4	5229	4	US-09-986-596C-16
21	256.4	4.6	259	4	US-09-016-434-402
22	238.2	4.3	44848	4	US-09-435-739-42
23	238.2	4.3	44848	4	US-09-988-113-42
24	233.4	4.2	460	4	US-09-635-872A-12
25	233.4	4.2	460	4	US-09-636-077A-12
26	233.4	4.2	460	4	US-09-636-060C-12
27	233.4	4.2	460	4	US-09-986-552-12

ALIGNMENTS

RESULT 1

US-09-635-872A-4

; Sequence 4, Application US/09635872A

; Patent No. 6534300

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE

; FILE REFERENCE: 195613US0

; CURRENT APPLICATION NUMBER: US/09/635,872A

; CURRENT FILING DATE: 2000-09-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 5597

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-635-872A-4

Query Match 100.0%; Score 5597; DB 4; Length 5597;

Best Local Similarity 100.0%; Pred No. 0;

Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

28 233.4 4.2 460 4 US-09-636-596C-12 Sequence 12, Appl
29 232 4.1 235 4 US-09-513-999C-23302 Sequence 23302, A
30 221.2 4.0 17327 1 US-07-906-871-15 Sequence 15, Appl
31 221 3.9 55298 4 US-09-491-356C-1 Sequence 1, Appl
32 217.2 3.9 3470 4 US-09-486-147-2 Sequence 2, Appl
33 217.2 3.9 128779 4 US-09-497-855A-38 Sequence 38, Appl
34 216.4 3.9 851 4 US-09-495-050A-138 Sequence 138, Appl
35 213.8 3.8 53332 4 US-09-801-861-3 Sequence 3, Appl
36 213.8 3.8 53332 4 US-10-224-562-3 Sequence 3, Appl
37 213.4 3.8 168174 4 US-10-071-411A-63 Sequence 63, Appl
38 213.4 3.8 168273 4 US-10-071-411A-2 Sequence 2, Appl
39 212.8 3.8 319608 4 US-09-539-333D-1 Sequence 1, Appl
40 212.2 3.8 331 4 US-09-513-999C-18965 Sequence 18965, A
41 212 3.8 99500 3 US-09-798-096-10 Sequence 10, Appl
42 211.8 3.8 1926 3 US-09-117-250-4 Sequence 4, Appl
43 211.6 3.8 713 4 US-09-513-999C-14742 Sequence 14742, A
44 211.6 3.8 4741 1 US-07-695-472B-4 Sequence 4, Appl
45 211.6 3.8 4741 4 US-09-106-375-4 Sequence 4, Appl

```

Db 361 AGAATCGGCTTTGTCTGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGCACAG 420
Qy 421 ATCTTGAATCTACTGAAGGAACCTACAGCAGGTCAGAGAACAGATGGAGGAGGACAGAAAG 480
Db 421 ATCTTGAATCTACTGAAGGAACCTACAGCAGGTCAGAGAACAGATGGAGGAGGACAGAAAG 480
Qy 481 CAATGAGAGAAATCCTTGGGAAAAACAACCGAACCTACTAAGAACAGTGAAGACAGT 540
Db 481 CAATGAGAGAAATCCTTGGGAAAAACAACCGAACCTACTAAGAACAGTGAAGACAGT 540
Qy 541 TAGAGTGTCTTAAACACACTGCAATTAAGGTGCCAATGCTTGCTGACCTGGACCGACCTGC 600
Db 541 TAGAGTGTCTTAAACACACTGCAATTAAGGTGCCAATGCTTGCTGACCTGGACCGACCTGC 600
Qy 601 CAGCCAAATCACCCTGAAGGACGTGCCATCTCTTTATCCCTCTTTTCAATCTGCCAGTG 660
Db 601 CAGCCAAATCACCCTGAAGGACGTGCCATCTCTTTATCCCTCTTTTCAATCTGCCAGTG 660
Qy 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTTGTTGTTTG 720
Db 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTTGTTGTTTG 720
Qy 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGACAGA 780
Db 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGACAGA 780
Qy 781 CAGTATGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGATAGTCTAATGCAAG 840
Db 781 CAGTATGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGATAGTCTAATGCAAG 840
Qy 841 ATTTGGCTTTCTCAGTGGATTTCCACCACATTCAGAGGAAACAAATCAACTAAAAACAA 900
Db 841 ATTTGGCTTTCTCAGTGGATTTCCACCACATTCAGAGGAAACAAATCAACTAAAAACAA 900
Qy 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTGTAFTCAGAGGCCAGTG 960
Db 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTGTAFTCAGAGGCCAGTG 960
Qy 961 TAGCGCTCTTAAACTGAATTAACCCAGGATTTTCAAGAAATGAATTAAGCAAACTAAGA 1020
Db 961 TAGCGCTCTTAAACTGAATTAACCCAGGATTTTCAAGAAATGAATTAAGCAAACTAAGA 1020
Qy 1021 AGAATGACCATTCAGTGGGAAAGAACTGACCAATAGTCTGATATTTATGCGATC 1080
Db 1021 AGAATGACCATTCAGTGGGAAAGAACTGACCAATAGTCTGATATTTATGCGATC 1080
Qy 1081 TGAGGCCATCAGCGAGTCTAAGCAGATGAGACATCTCTGCCAGTCGTTTGAAGATA 1140
Db 1081 TGAGGCCATCAGCGAGTCTAAGCAGATGAGACATCTCTGCCAGTCGTTTGAAGATA 1140
Qy 1141 ACGAAGAACTCAGGTACTCTATGCGATCTATCGAGAGGCAATGACCAATGGTTTCGGAATA 1200
Db 1141 ACGAAGAACTCAGGTACTCTATGCGATCTATCGAGAGGCAATGACCAATGGTTTCGGAATA 1200
Qy 1201 TTTTCATTTGCAACCGGAGATTCATCTCTGGCTGGAACCTTGACAATCTCTCGAGTGA 1260
Db 1201 TTTTCATTTGCAACCGGAGATTCATCTCTGGCTGGAACCTTGACAATCTCTCGAGTGA 1260
Qy 1261 CAATAGTAACACACCGAGATTTTTCGAAATTTGACCACTTGCCTTACCTTTAGTTTAC 1320
Db 1261 CAATAGTAACACACCGAGATTTTTCGAAATTTGACCACTTGCCTTACCTTTAGTTTAC 1320
Qy 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAGAGTTTATTAACCTAA 1380
Db 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAGAGTTTATTAACCTAA 1380
Qy 1381 ATGATGATGTCATGTTTGGGAGGATGCTGCGCAGATGATTTTACAGTCACCTCCAAAG 1440
Db 1381 ATGATGATGTCATGTTTGGGAGGATGCTGCGCAGATGATTTTACAGTCACCTCCAAAG 1440
Qy 1441 GCCAAGAGGTTTATTTGACATGCGCTGTGCCAACTGTGCGGAGGCTGCCAGGTTCTT 1500
Db 1441 GCCAAGAGGTTTATTTGACATGCGCTGTGCCAACTGTGCGGAGGCTGCCAGGTTCTT 1500

Qy 1501 GGATTAAGGATGGCTATTGTGCAAGGCTTGTAAATAATTCAGCCTGCGATTGGGATGGT 1560
Db 1501 GGATTAAGGATGGCTATTGTGCAAGGCTTGTAAATAATTCAGCCTGCGATTGGGATGGT 1560
Qy 1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCGCTATATTGCAAGGAGTGGAGGTAAGTGGGA 1620
Db 1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCGCTATATTGCAAGGAGTGGAGGTAAGTGGGA 1620
Qy 1621 GTATTGGAGTGGACACCCCTGGCAGTTTGTGGAGGAATAAAACAGTGTCTCTTACTGTA 1680
Db 1621 GTATTGGAGTGGACACCCCTGGCAGTTTGTGGAGGAATAAAACAGTGTCTCTTACTGTA 1680
Qy 1681 ATCAGGGATGTCGAAATCTCTGGCTGCTGATATAAGTCTGTGACCAAGCATGCAATGTCT 1740
Db 1681 ATCAGGGATGTCGAAATCTCTGGCTGCTGATATAAGTCTGTGACCAAGCATGCAATGTCT 1740
Qy 1741 TGTCTGTGGTGTGATGCTGCGGACTGTGGGCAAGATCAITTTTCATGAATGTATAAAG 1800
Db 1741 TGTCTGTGGTGTGATGCTGCGGACTGTGGGCAAGATCAITTTTCATGAATGTATAAAG 1800
Qy 1801 TGATCTCTTCTCCAAACCCAGACTCACTATATTATTTCAAAAAGTGAATGCTGCTTATT 1860
Db 1801 TGATCTCTTCTCCAAACCCAGACTCACTATATTATTTCAAAAAGTGAATGCTGCTTATT 1860
Qy 1861 TCAGCTTTGCAAGTAGTCCAAAGAGGAGTGAAGTGCCTATAGTACATCCAAATAA 1920
Db 1861 TCAGCTTTGCAAGTAGTCCAAAGAGGAGTGAAGTGCCTATAGTACATCCAAATAA 1920
Qy 1921 TTCGACATGCTTCTATTGCCAAACAGTGGAAACCATCCACTCATATATGACAGTGGAA 1980
Db 1921 TTCGACATGCTTCTATTGCCAAACAGTGGAAACCATCCACTCATATATGACAGTGGAA 1980
Qy 1981 TGAATGCCACCAACATACTTTTAACTCTCAGCTTTTCAAAATACAAACGATGAAGTTC 2040
Db 1981 TGAATGCCACCAACATACTTTTAACTCTCAGCTTTTCAAAATACAAACGATGAAGTTC 2040
Qy 2041 AAATCCAGATTAACAGTGGAGTGGACCAAGGGAGGACCAAACTGAAATTTACGGCCC 2100
Db 2041 AAATCCAGATTAACAGTGGAGTGGACCAAGGGAGGACCAAACTGAAATTTACGGCCC 2100
Qy 2101 AGAAGGTTACGAAATTTAGTGTAGTCCATAACACTTTTCCAGAGCGGAAATCTTTT 2160
Db 2101 AGAAGGTTACGAAATTTAGTGTAGTCCATAACACTTTTCCAGAGCGGAAATCTTTT 2160
Qy 2161 TTGAGGATATTTCCAAAGAAACGCTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220
Db 2161 TTGAGGATATTTCCAAAGAAACGCTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220
Qy 2221 CAAGGAGAGCCAGGAGAGTGAATTTCCCTGGTAAATTTTCACTCTTCTCCAAAG 2280
Db 2221 CAAGGAGAGCCAGGAGAGTGAATTTCCCTGGTAAATTTTCACTCTTCTCCAAAG 2280
Qy 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCACTGGAACTGGAGACATCACTTTGA 2340
Db 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCACTGGAACTGGAGACATCACTTTGA 2340
Qy 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAITTTCTGATGAACCTCACAGCATG 2400
Db 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAITTTCTGATGAACCTCACAGCATG 2400
Qy 2401 CTAAATTAATAAATCAAGCTATAATAACAGATGAATAAATGACAGTTTGGTGGCTCCAC 2460
Db 2401 CTAAATTAATAAATCAAGCTATAATAACAGATGAATAAATGACAGTTTGGTGGCTCCAC 2460
Qy 2461 AGGAAAAACAGGTTTCATAAAAGCATCTTCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
Db 2461 AGGAAAAACAGGTTTCATAAAAGCATCTTCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
Qy 2521 AGAGTTGACCTTTCTGCACTGAGTGTAAAGTGAATGTCATGACAGGTCAGAAATC 2580
Db 2521 AGAGTTGACCTTTCTGCACTGAGTGTAAAGTGAATGTCATGACAGGTCAGAAATC 2580

QY 2581 CACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACAACCCCAAAAAACCA 2640
Db 2581 CACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACAACCCCAAAAAACCA 2640
QY 2641 TAGGCGGAAATGTGCAAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGCCGAGA 2700
Db 2641 TAGGCGGAAATGTGCAAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGCCGAGA 2700
QY 2701 TGACAAAAGAAAGAAATCAGGGAAGAAAGAAAGCAAGTAGAATGAGGAAATG 2760
Db 2701 TGACAAAAGAAAGAAATCAGGGAAGAAAGAAAGCAAGTAGAATGAGGAAATG 2760
QY 2761 CTGAAAATCACAATAGGCGTACTGAAAGTGTACTTTGGAAAGAAAGCTGCAAGATTTACACAG 2820
Db 2761 CTGAAAATCACAATAGGCGTACTGAAAGTGTACTTTGGAAAGAAAGCTGCAAGATTTACACAG 2820
QY 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAGTATTTCCAGATCTTCTCGACG 2880
Db 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAGTATTTCCAGATCTTCTCGACG 2880
QY 2881 AAGAAAGATCAATTTGAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGAGGC 2940
Db 2881 AAGAAAGATCAATTTGAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGAGGC 2940
QY 2941 AACTAAAAGATACATTTGGAGATCCCTCAGATATGTAAATTAATAATAGCAAGT 3000
Db 2941 AACTAAAAGATACATTTGGAGATCCCTCAGATATGTAAATTAATAATAGCAAGT 3000
QY 3001 TTGGATTCACATCGGGAAGTCCCTGCTCAGATGCTCAGATGATTCACCGGATGTGTA 3060
Db 3001 TTGGATTCACATCGGGAAGTCCCTGCTCAGATGCTCAGATGATTCACCGGATGTGTA 3060
QY 3061 TGCAAGAACTCAAGATATGTTCCCTGAAGATTTGCAAGACGTCATTTTCAAAAAGTGC 3120
Db 3061 TGCAAGAACTCAAGATATGTTCCCTGAAGATTTGCAAGACGTCATTTTCAAAAAGTGC 3120
QY 3121 GCAATTCAGAGATATGAGTTGCTTCTTATTTTATATCTCATGATGTCAGTGC 3180
Db 3121 GCAATTCAGAGATATGAGTTGCTTCTTATTTTATATCTCATGATGTCAGTGC 3180
QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTTGT 3240
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTTGT 3240
QY 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAAGTTGCAAG 3300
Db 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAAGTTGCAAG 3300
QY 3301 ATTTGACAGTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360
Db 3301 ATTTGACAGTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360
QY 3361 CGCAGCTAAATATTTCCAACTCAGGAATCTTATGATCCCAACCTGCGACCGG 3420
Db 3361 CGCAGCTAAATATTTCCAACTCAGGAATCTTATGATCCCAACCTGCGACCGG 3420
QY 3421 TCACATAAAGTCTAGTACAACTCTAACCAGTAACTGACAAAATCCACAAGCATATA 3480
Db 3421 TCACATAAAGTCTAGTACAACTCTAACCAGTAACTGACAAAATCCACAAGCATATA 3480
QY 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
Db 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
QY 3541 TTGCTACCAAGTCTCATGTTGTTGGCAGTTGGATGACATAAGAAAACCCCTAGGA 3600
Db 3541 TTGCTACCAAGTCTCATGTTGTTGGCAGTTGGATGACATAAGAAAACCCCTAGGA 3600
QY 3601 AGTTTGTGCTGCTGAATGACAACTTGACCAATCAATAAGATGCTCAGACAGTGAAG 3660
Db 3601 AGTTTGTGCTGCTGAATGACAACTTGACCAATCAATAAGATGCTCAGACAGTGAAG 3660
QY 3661 CTGTTCTCAGGAGCTTCTATGAATCAATGTTCCCATACCTTTCCCAATTTGAACTGCCAA 3720

Db 3661 CTGTTCTCAGGAGCTTCTATGAATCAATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720
QY 3721 GAGAGTATCGAAACCGTTTCCCTTCATATGATGAGCTGAGAGTCCAGAGCTTATCGAG 3780
Db 3721 GAGAGTATCGAAACCGTTTCCCTTCATATGATGAGCTGAGAGTCCAGAGCTTATCGAG 3780
QY 3781 ACAAATTTGAAGTTTTCGACCCCATTTGTTACTAGCAACATTTGTTACTATATTTCT 3840
Db 3781 ACAAATTTGAAGTTTTCGACCCCATTTGTTACTAGCAACATTTGTTACTATATTTCT 3840
QY 3841 CATTTTTCGCTGAGAGTTAAATTTGCACTTAAGCGGAAGATATTTCCCAAGAGAGATAC 3900
Db 3841 CATTTTTCGCTGAGAGTTAAATTTGCACTTAAGCGGAAGATATTTCCCAAGAGAGATAC 3900
QY 3901 ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCAATTTGAAAACCACTA 3960
Db 3901 ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCAATTTGAAAACCACTA 3960
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTGTGATGTGAT 4020
Db 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTGTGATGTGAT 4020
QY 4021 GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCAGTCTGTTTGTGGCATGA 4080
Db 4021 GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCAGTCTGTTTGTGGCATGA 4080
QY 4081 TATAGCCCACTGACTAGGAATTTAAACCAACCCACTGAAAACCTTGTGTGCGAGCAGC 4140
Db 4081 TATAGCCCACTGACTAGGAATTTAAACCAACCCACTGAAAACCTTGTGTGCGAGCAGC 4140
QY 4141 TCTGAACATGATTTACTTTTAAAGATTTGCTCATGAGCTGTCATCTTTTATAAAA 4200
Db 4141 TCTGAACATGATTTACTTTTAAAGATTTGCTCATGAGCTGTCATCTTTTATAAAA 4200
QY 4201 GGTCTACTGACAAAGACAGCTGTTAAATTTCCACAGCAATCAATGACAGTAACTTTAT 4260
Db 4201 GGTCTACTGACAAAGACAGCTGTTAAATTTCCACAGCAATCAATGACAGTAACTTTAT 4260
QY 4261 TAGGAGAGCCTATGCCAGCTGGAGTGTGTAAGAGGCTCCAGCTTTGCAATCCAA 4320
Db 4261 TAGGAGAGCCTATGCCAGCTGGAGTGTGTAAGAGGCTCCAGCTTTGCAATCCAA 4320
QY 4321 AGCCTTTGCTAAAGTTTTCACATTTTTCATTTTCCCATTTTAAAGTAGTTACTA 4380
Db 4321 AGCCTTTGCTAAAGTTTTCACATTTTTCATTTTCCCATTTTAAAGTAGTTACTA 4380
QY 4381 AGTTAACTAGTTATTTCTGCTTCTGATATAACGAATTTGGATGCTTAAACCTTATTTA 4440
Db 4381 AGTTAACTAGTTATTTCTGCTTCTGATATAACGAATTTGGATGCTTAAACCTTATTTA 4440
QY 4441 TAGATGTTTAAATAATGACAGCAATATCACCTCTTATTTGACAATACCTTAATTTAG 4500
Db 4441 TAGATGTTTAAATAATGACAGCAATATCACCTCTTATTTGACAATACCTTAATTTAG 4500
QY 4501 TTTTATTAATTTAAGACTGTAATTTGCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560
Db 4501 TTTTATTAATTTAAGACTGTAATTTGCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560
QY 4561 ATTGACATCTGAAATGCTTTGTAATTTAGCTTTAGCTTTAGCTTTAGCTTTAGCTTTA 4620
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTAGCTTTAGCTTTAGCTTTAGCTTTA 4620
QY 4621 CGTGCAGCTCTAATTTCAACAGGCTAGGTTAGTACTTACTTACAGATGTAATTTATGTTT 4680
Db 4621 CGTGCAGCTCTAATTTCAACAGGCTAGGTTAGTACTTACTTACAGATGTAATTTATGTTT 4680
QY 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGATGAGTAGTGTGATGGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGATGAGTAGTGTGATGGCA 4740
QY 4741 CTGCAATTTACAGTGGTGTCTTTTAAATACCTCATTGTTATTTCCAGTAGCTATCTCT 4800

Db	3121	GCATCTCTGAGGATATGAGTTGGCTTCTCTTATTTATTTATCTCATGAGTCAGTGC	3180
Qy	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTCTTGT	3240
Db	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTCTTGT	3240
Qy	3241	CTGACAGAGAAATCCGAACACTGGCTACCGAATTCAGAACTCCGCTTAAGTTTGAGG	3300
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACCGAATTCAGAACTCCGCTTAAGTTTGAGG	3300
Qy	3301	ATTTGACAGGTTCTGGAAACACATGCTTAATAATTCGTCAAAATGCTTCCTGCTGATATCA	3360
Db	3301	ATTTGACAGGTTCTGGAAACACATGCTTAATAATTCGTCAAAATGCTTCCTGCTGATATCA	3360
Qy	3361	CGAGCTAAATTAATTTCCCAACTCCAGGAATCCTACTATGATCCCAACTGCCACCGG	3420
Db	3361	CGAGCTAAATTAATTTCCCAACTCCAGGAATCCTACTATGATCCCAACTGCCACCGG	3420
Qy	3421	TCACATAAAGTCTAGTACAACTGTAAACAGTAATCTGACAAATCCAGAGCATATA	3480
Db	3421	TCACATAAAGTCTAGTACAACTGTAAACAGTAATCTGACAAATCCAGAGCATATA	3480
Qy	3481	AGGACAAAACAAATATAGTTTGAATATCATGGGAGAGAAATCGCTTTTAAATGA	3540
Db	3481	AGGACAAAACAAATATAGTTTGAATATCATGGGAGAGAAATCGCTTTTAAATGA	3540
Qy	3541	TTCTGACCAAGTTTCTCATGTTGGTGGCCAGTTGGATGACATAAGAAAAAACCCTAGGA	3600
Db	3541	TTCTGACCAAGTTTCTCATGTTGGTGGCCAGTTGGATGACATAAGAAAAAACCCTAGGA	3600
Qy	3601	AGTTGTTGCTGCTGAATCAACAACTGACCAATCAATAAGATGCTCAGACAGTGAAG	3660
Db	3601	AGTTGTTGCTGCTGAATCAACAACTGACCAATCAATAAGATGCTCAGACAGTGAAG	3660
Qy	3661	CTGTTCTCAGGACTCTATGAATCCATGTTTCCCACTACCTTCCCAATTTGAACTGCCAA	3720
Db	3661	CTGTTCTCAGGACTCTATGAATCCATGTTTCCCACTACCTTCCCAATTTGAACTGCCAA	3720
Qy	3721	GAGAGTATCGAAACCGTTTCTCATATGATGATGAGTGCAGGATGAGGCTTATCGAG	3780
Db	3721	GAGAGTATCGAAACCGTTTCTCATATGATGATGAGTGCAGGATGAGGCTTATCGAG	3780
Qy	3781	ACAAATTTGAAGTTTGGACCCATTTGTTTCCCACTACCTTCCCAATTTGAACTGCCAA	3840
Db	3781	ACAAATTTGAAGTTTGGACCCATTTGTTTCCCACTACCTTCCCAATTTGAACTGCCAA	3840
Qy	3841	CATTTTGTGCTGAGCAGTAAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC	3900
Db	3841	CATTTTGTGCTGAGCAGTAAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC	3900
Qy	3901	ACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAAACCATCTA	3960
Db	3901	ACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAAACCATCTA	3960
Qy	3961	CCTCAGCAATTTACTGAGCAGTAAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC	4020
Db	3961	CCTCAGCAATTTACTGAGCAGTAAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC	4020
Qy	4021	GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGGCCATGAA	4080
Db	4021	GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGGCCATGAA	4080
Qy	4081	TATAGCCCACTGACTAGAAATTTTAAACCAACCCACTGTAATAAATTTTATAAAAA	4140
Db	4081	TATAGCCCACTGACTAGAAATTTTAAACCAACCCACTGTAATAAATTTTATAAAAA	4140
Qy	4141	TCTGAATCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCTGCTTTTATAAAAA	4200
Db	4141	TCTGAATCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCTGCTTTTATAAAAA	4200
Qy	4201	GGCTCCTGACAGAGACAGTGTAAATTTCCACAGCAATCAATGCAGACTAATTTAT	4260

Db	4201	GGCTCCTGACAGAGACAGCTGTAAATTTCCACAGCAATCAATTCAGACTAATTTAT	4260
Qy	4261	TAGGAGAACCTTATGCGAGCTGGAGTGAATCTCTAAGAGGCTCCAGTCTTTGCACTTCAA	4320
Db	4261	TAGGAGAACCTTATGCGAGCTGGAGTGAATCTCTAAGAGGCTCCAGTCTTTGCACTTCAA	4320
Qy	4321	AGCCTTTTCTGCTAAAGTTTGGACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	4380
Db	4321	AGCCTTTTCTGCTAAAGTTTGGACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	4380
Qy	4381	AGTTAACTAGTTTATTTCTTCTGCTTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTA	4440
Db	4381	AGTTAACTAGTTTATTTCTTCTGCTTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTA	4440
Qy	4441	TAGATGTTTATTTTAAATTAATTCAGCAATATCACTCTTATTTGACAAATCACTAAATTTATG	4500
Db	4441	TAGATGTTTATTTTAAATTAATTCAGCAATATCACTCTTATTTGACAAATCACTAAATTTATG	4500
Qy	4501	TTTTTATTAATTTTAAAGTCTGTAATGCTTTAAACCACTACTACTGAGAGCTCAATG	4560
Db	4501	TTTTTATTAATTTTAAAGTCTGTAATGCTTTAAACCACTACTACTGAGAGCTCAATG	4560
Qy	4561	ATTGACATCTGAAATGCTTTTGTAAATTTTGAATTTTGAATTTTGAATTTGCTATGATTTCA	4620
Db	4561	ATTGACATCTGAAATGCTTTTGTAAATTTTGAATTTTGAATTTTGAATTTGCTATGATTTCA	4620
Qy	4621	CGTGCAGGCTTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT	4680
Db	4621	CGTGCAGGCTTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT	4680
Qy	4681	TGGAATGTACATATTTCAACAGAGTGCCTCAITTTAGAAATGAGTAGTCTGATGSCA	4740
Db	4681	TGGAATGTACATATTTCAACAGAGTGCCTCAITTTAGAAATGAGTAGTCTGATGSCA	4740
Qy	4741	CTGGCACAATTAAGTGTGCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	4800
Db	4741	CTGGCACAATTAAGTGTGCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	4800
Qy	4801	CTCAGTTGTTTTCATAGAACAGAGGCGCAGCAACTTTCTTTGTAAGGCTGGTGTAGT	4860
Db	4801	CTCAGTTGTTTTCATAGAACAGAGGCGCAGCAACTTTCTTTGTAAGGCTGGTGTAGT	4860
Qy	4861	AAATTTATTCAGGCGCACCTGCTCTTCTCATACATTTCTTCTGCTGTTTGTAGTTTGT	4920
Db	4861	AAATTTATTCAGGCGCACCTGCTCTTCTCATACATTTCTTCTGCTGTTTGTAGTTTGT	4920
Qy	4921	TTTTTTTCAAAACACCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4980
Db	4921	TTTTTTTCAAAACACCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4980
Qy	4981	ACTGCCACCCAGCAGATGTGACCCCTCAGGCGCATCATTTGCCAATCACTGAGAAATTTT	5040
Db	4981	ACTGCCACCCAGCAGATGTGACCCCTCAGGCGCATCATTTGCCAATCACTGAGAAATTTT	5040
Qy	5041	TTGTTTGTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT	5100
Db	5041	TTGTTTGTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT	5100
Qy	5101	TGCACTGGCGCAATCTCAGCTCACTGCAACCTCCGCGCTCCCGGGTTCAAGCAGTTCTGTC	5160
Db	5101	TGCACTGGCGCAATCTCAGCTCACTGCAACCTCCGCGCTCCCGGGTTCAAGCAGTTCTGTC	5160
Qy	5161	TCAGCCTTCTGAGTAGCTGGGACTACAGGTGCTGCAACCAACCTCTCTCTCTCTCTCTCT	5220
Db	5161	TCAGCCTTCTGAGTAGCTGGGACTACAGGTGCTGCAACCAACCTCTCTCTCTCTCTCTCT	5220
Qy	5221	TTTTTGTAGTAGAGAGGGGGTTTCCACCATTTTGTGTCAGGCTTATCTGAACTCTGACCTC	5280
Db	5221	TTTTTGTAGTAGAGAGGGGGTTTCCACCATTTTGTGTCAGGCTTATCTTGAACCTCTGACCTC	5280
Qy	5281	AGGTGATCCACTGCTCTGCTCTCCCAAGTGTGAGATTACAGGCATATAAGCCAGTGCAC	5340
Db	5281	AGGTGATCCACTGCTCTGCTCTCCCAAGTGTGAGATTACAGGCATATAAGCCAGTGCAC	5340

QY	1501	GGATTAGGATGGCTATTGTGACAAAGGCTTGTAATAATTCAGCTCGATGGGATGGTG	1560
Db	1501	GGATTAGGATGGCTATTGTGACAAAGGCTTGTAATAATTCAGCTCGATGGGATGGTG	1560
QY	1561	GGATTGCTCTGGAACAGTGGAGGAGTCCCTATATTTGAGGAGGTTGGAGTACTGGGA	1620
Db	1561	GGATTGCTCTGGAACAGTGGAGGAGTCCCTATATTTGAGGAGGTTGGAGTACTGGGA	1620
QY	1621	GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAAGTGTCTTCTTACTGTA	1680
Db	1621	GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAAGTGTCTTCTTACTGTA	1680
QY	1681	ATCAGGATGTCGGAATTTCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT	1740
Db	1681	ATCAGGATGTCGGAATTTCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT	1740
QY	1741	TGTCCTGTGGTTTGCATCTGGGACTGTGGGCAAGATCATTTTCATGAATTTGTATAAG	1800
Db	1741	TGTCCTGTGGTTTGCATCTGGGACTGTGGGCAAGATCATTTTCATGAATTTGTATAAG	1800
QY	1801	TGATCCTTCTCCAAACAGACTCACTATATTTTCCAAAGGTTGAATGCTGCTCTTATT	1860
Db	1801	TGATCCTTCTCCAAACAGACTCACTATATTTTCCAAAGGTTGAATGCTGCTCTTATT	1860
QY	1861	TCAGCTTTCAGAGTAGCCAAAGAGAGTTGAAGTGCTATAGTGACAATCCCAATAA	1920
Db	1861	TCAGCTTTCAGAGTAGCCAAAGAGAGTTGAAGTGCTATAGTGACAATCCCAATAA	1920
QY	1921	TTCGACATGCTTCTATTGCGCAAGTGGAAACCATCCACTCATATGACACAGTGGAA	1980
Db	1921	TTCGACATGCTTCTATTGCGCAAGTGGAAACCATCCACTCATATGACACAGTGGAA	1980
QY	1981	TGAATGCCACACAAATACATTTTAACTCAGTTTCAAATACAAAGATGAAGTTCA	2040
Db	1981	TGAATGCCACACAAATACATTTTAACTCAGTTTCAAATACAAAGATGAAGTTCA	2040
QY	2041	AAATGTCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC	2100
Db	2041	AAATGTCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC	2100
QY	2101	AGAGGTTAGCAAAATTTAGTTAGTCCATPAAACACTTTCTCAGAGGCGGAAATCTTTT	2160
Db	2101	AGAGGTTAGCAAAATTTAGTTAGTCCATPAAACACTTTCTCAGAGGCGGAAATCTTTT	2160
QY	2161	TTGAGGATTTCCCAAGAAAAACGTTCCCGAAGTTTAAAGACATGATTTAACTCAA	2220
Db	2161	TTGAGGATTTCCCAAGAAAAACGTTCCCGAAGTTTAAAGACATGATTTAACTCAA	2220
QY	2221	CAAGGAGGCCAGGAGAGTGAATAATTCCTTGGTAAATTTTCACTCCTTCCAAAAG	2280
Db	2221	CAAGGAGGCCAGGAGAGTGAATAATTCCTTGGTAAATTTTCACTCCTTCCAAAAG	2280
QY	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGACATGGAGACATCACTTTGA	2340
Db	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGACATGGAGACATCACTTTGA	2340
QY	2341	AAGGATCAATTTGTCCAAAGTCAGCTTGTGAGATCATTTTGTGATGAATTCACAGATG	2400
Db	2341	AAGGATCAATTTGTCCAAAGTCAGCTTGTGAGATCATTTTGTGATGAATTCACAGATG	2400
QY	2401	CTAAAAATAAAAATCAAGCTTAATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC	2460
Db	2401	CTAAAAATAAAAATCAAGCTTAATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC	2460
QY	2461	AGGAAAAACAGTTTCAATAAAGCATCTTGGCCAAACAGTTAGGAGTGTCTGAAAGATTGC	2520
Db	2461	AGGAAAAACAGTTTCAATAAAGCATCTTGGCCAAACAGTTAGGAGTGTCTGAAAGATTGC	2520
QY	2521	AGAGGTTGACTTTTCTCAGTGTAGTGAAGTGAATGTTGATGACAGGCTCAGATC	2580
Db	2521	AGAGGTTGACTTTTCTCAGTGTAGTGAAGTGAATGTTGATGACAGGCTCAGATC	2580
QY	2581	CACCCCTGGACTTTGGAGACCAACAGATTTAGAGTGGAAATCTCACACCCAAAAACCA	2640

Db	2581	CACCCCTGGACTTTGGAGACCAACAGATTTAGAGTGGAAATCTCACACCCAAAAACCA	2640
QY	2641	TAGGCGGAAATCTGACAAAAGAAAAAGCCCATCTCTGATTGTTCCACTGGAAGCCGAGA	2700
Db	2641	TAGGCGGAAATCTGACAAAAGAAAAAGCCCATCTCTGATTGTTCCACTGGAAGCCGAGA	2700
QY	2701	TCACAAAAGAAAATTCACAGGAAAAGAAAAGAACAGTAGAATGAGGAAAAATG	2760
Db	2701	TCACAAAAGAAAATTCACAGGAAAAGAAAAGAACAGTAGAATGAGGAAAAATG	2760
QY	2761	CTGAAAATCACATAGCGCTTACTGAAGTGTACTTTGGAGAAAGCTGAGCATTCACAG	2820
Db	2761	CTGAAAATCACATAGCGCTTACTGAAGTGTACTTTGGAGAAAGCTGAGCATTCACAG	2820
QY	2821	ATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTCCAGATCTTCTCGAGC	2880
Db	2821	ATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTCCAGATCTTCTCGAGC	2880
QY	2881	AAGAAGATCATTTGAAGACAAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC	2940
Db	2881	AAGAAGATCATTTGAAGACAAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC	2940
QY	2941	AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT	3000
Db	2941	AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT	3000
QY	3001	TTGGATTCAATTCGGGAAAGTCCCTGCTCATGCTCTCAATGATGACCGGATTTGTTA	3060
Db	3001	TTGGATTCAATTCGGGAAAGTCCCTGCTCATGCTCTCAATGATGACCGGATTTGTTA	3060
QY	3061	TCCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGCAAGACGTCATTTCAAAAGTGC	3120
Db	3061	TCCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGCAAGACGTCATTTCAAAAGTGC	3120
QY	3121	GCCATTCTGAGGATGAGTGTTCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC	3180
Db	3121	GCCATTCTGAGGATGAGTGTTCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC	3180
QY	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT	3240
Db	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT	3240
QY	3241	CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTCCGTTAAATTTGCAGG	3300
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTCCGTTAAATTTGCAGG	3300
QY	3301	ATTTGACAGGCTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTGCTGATATCA	3360
Db	3301	ATTTGACAGGCTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTGCTGATATCA	3360
QY	3361	CGCAGTAAATAATATTCACCAACTCAGGAATCTTACTATGATCCCAACTGCCACCGG	3420
Db	3361	CGCAGTAAATAATATTCACCAACTCAGGAATCTTACTATGATCCCAACTGCCACCGG	3420
QY	3421	TCCTAAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCATATA	3480
Db	3421	TCCTAAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCATATA	3480
QY	3481	AGGACAAAACAAATATAGTTTGAATCATGGGAGAGAAATCCCTTTTAAATATGA	3540
Db	3481	AGGACAAAACAAATATAGTTTGAATCATGGGAGAGAAATCCCTTTTAAATATGA	3540
QY	3541	TTGCTACCAAGTCTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAACCTTAGGA	3600
Db	3541	TTGCTACCAAGTCTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAACCTTAGGA	3600
QY	3601	AGTTTGTGCTGATGACACATTTGACCAATCATTAAGATGCTCAGACAGTGAAGG	3660
Db	3601	AGTTTGTGCTGATGACACATTTGACCAATCATTAAGATGCTCAGACAGTGAAGG	3660
QY	3661	CTGTTCTCAGGACTTCTTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCGCAA	3720

Db 3661 CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCCAA 3720
Qy 3721 GAGATATCGAAACCGTTTCTTATATGCAATGAGCTGCAGGAATGAGGGCTTATCGAG 3780
Db 3721 GAGATATCGAAACCGTTTCTTATATGCAATGAGCTGCAGGAATGAGGGCTTATCGAG 3780
Qy 3781 ACAAAATGAAGTTTGGACCAATGCTGATAGCAAACTGATATATGTTTACTATATCT 3840
Db 3781 ACAAAATGAAGTTTGGACCAATGCTGATAGCAAACTGATATATGTTTACTATATCT 3840
Qy 3841 CATTTTCTGAGCAGTTAAATGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC 3900
Db 3841 CATTTTCTGAGCAGTTAAATGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC 3900
Qy 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGTATCTCATTTGAAACCACTTA 3960
Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGTATCTCATTTGAAACCACTTA 3960
Qy 3961 CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTCTTCTGATGAT 4020
Db 3961 CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTCTTCTGATGAT 4020
Qy 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGTGCGATGAA 4080
Db 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGTGCGATGAA 4080
Qy 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACTTGTGCTGCGAGCAGC 4140
Db 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACTTGTGCTGCGAGCAGC 4140
Qy 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGTCATCTTTTATAAAA 4200
Db 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGTCATCTTTTATAAAA 4200
Qy 4201 GGCTCAGTGACAGAGCAGCTGTTAAATTTCCCAAGCAATCAATGCACTACTTAT 4260
Db 4201 GGCTCAGTGACAGAGCAGCTGTTAAATTTCCCAAGCAATCAATGCACTACTTAT 4260
Qy 4261 TAGGAGAGCCCTATGCCAGCTGGAGTGATTTGCTAAGAGGCTCCAGTCTTGCATTCCTCA 4320
Db 4261 TAGGAGAGCCCTATGCCAGCTGGAGTGATTTGCTAAGAGGCTCCAGTCTTGCATTCCTCA 4320
Qy 4321 AGCCTTTTGTAAAGTTTGCATTTTATTTTTCATTTCCATTTTAAAGTAGTACTA 4380
Db 4321 AGCCTTTTGTAAAGTTTGCATTTTATTTTTCATTTCCATTTTAAAGTAGTACTA 4380
Qy 4381 AGTTAACTAGTTATTTCTTCTGAGTAAACGAATTTGGATGCTTAAACCTTATTTTA 4440
Db 4381 AGTTAACTAGTTATTTCTTCTGAGTAAACGAATTTGGATGCTTAAACCTTATTTTA 4440
Qy 4441 TAGATGTTATTTAAATAGCAGCAATATCACCTTTATTCACCAATACCTTAAATATGAG 4500
Db 4441 TAGATGTTATTTAAATAGCAGCAATATCACCTTTATTCACCAATACCTTAAATATGAG 4500
Qy 4501 TTTTATTAATTTTAAAGCTGTAATGCTTTAAACCACTTAACTACTGAGAGCTCAATG 4560
Db 4501 TTTTATTAATTTTAAAGCTGTAATGCTTTAAACCACTTAACTACTGAGAGCTCAATG 4560
Qy 4561 ATTGACATCTGAATGCTTTGTAATTTATTTGCTTCCAGCCCTTAAAGATGCTATGATTCA 4620
Db 4561 ATTGACATCTGAATGCTTTGTAATTTATTTGCTTCCAGCCCTTAAAGATGCTATGATTCA 4620
Qy 4621 CGTGAGCTTAATTTTCAAGCTGAGATTTAGTACTTACTTACAGATGTAATTTATGTT 4680
Db 4621 CGTGAGCTTAATTTTCAAGCTGAGATTTAGTACTTACTTACAGATGTAATTTATGTT 4680
Qy 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTGAAGATGAGTAGTCTGATGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTGAAGATGAGTAGTCTGATGCA 4740
Qy 4741 CTGGCACAATTACAGTGGTCTGTTTAAATCTCATTTGATATTTCCAGTAGCTATCTCT 4800
Db 4741 CTGGCACAATTACAGTGGTCTGTTTAAATCTCATTTGATATTTCCAGTAGCTATCTCT 4800

Qy 4801 CTCAGTTGGTTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860
Db 4801 CTCAGTTGGTTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860
Qy 4861 AAATTAATGAGGCCACCTGTTGTTTGTATACATCTTCTGCTGTTGTTAGTTGT 4920
Db 4861 AAATTAATGAGGCCACCTGTTGTTTGTATACATCTTCTGCTGTTGTTAGTTGT 4920
Qy 4921 TTTTTCATCAAAACCTCTTAAATATGTAATAAATGTAATAAATGTAATAAATGTAATAA 4980
Db 4921 TTTTTCATCAAAACCTCTTAAATATGTAATAAATGTAATAAATGTAATAAATGTAATAA 4980
Qy 4981 ACTGCCACAGCCAGATGTCAGCTCAGCCCTCAGCCCATCATTTGCCAATCAGTGAATTTT 5040
Db 4981 ACTGCCACAGCCAGATGTCAGCTCAGCCCTCAGCCCATCATTTGCCAATCAGTGAATTTT 5040
Qy 5041 TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTTGCCAGCTGGAG 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTTGCCAGCTGGAG 5100
Qy 5101 TGCAGTGGGCAATCTCAGCTCAGCTCAGCCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 5160
Db 5101 TGCAGTGGGCAATCTCAGCTCAGCTCAGCCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 5160
Qy 5161 TCAGCCTTCTGAGTACCTGGGACTACAGGTGATGCCACACACACCTGCTTAATTTTGT 5220
Db 5161 TCAGCCTTCTGAGTACCTGGGACTACAGGTGATGCCACACACACCTGCTTAATTTTGT 5220
Qy 5221 TTTTATGATAGACGGGGTTTCCACATATTTGCTCAGGCTTATCTTTGAACTCTGACCTC 5280
Db 5221 TTTTATGATAGACGGGGTTTCCACATATTTGCTCAGGCTTATCTTTGAACTCTGACCTC 5280
Qy 5281 AGTGATCCACCTGCTGCTCCCAAGTCTGAGATTACAGGATTAAGCCAGTGCAC 5340
Db 5281 AGTGATCCACCTGCTGCTCCCAAGTCTGAGATTACAGGATTAAGCCAGTGCAC 5340
Qy 5341 CCAGCCGAGATTTATTTTATGATGTTTAAACCTTTGGCTCTAGCCATTTTATG 5400
Db 5341 CCAGCCGAGATTTATTTTATGATGTTTAAACCTTTGGCTCTAGCCATTTTATG 5400
Qy 5401 TCATAATCAATGATTTCTGAGACAGATTCATGAGTACTCTGACAGGTATTTTATG 5460
Db 5401 TCATAATCAATGATTTCTGAGACAGATTCATGAGTACTCTGACAGGTATTTTATG 5460
Qy 5461 ATCATGATCTCAACAATATTCCTCCCAATGTCATACATCTTTTGTACAAAGACTTGA 5520
Db 5461 ATCATGATCTCAACAATATTCCTCCCAATGTCATACATCTTTTGTACAAAGACTTGA 5520
Qy 5521 ATGTAATATCTGTTGTTGCTGTAAGAGTGTGTAATTTCAAAAATCTGAAATCTCATA 5580
Db 5521 ATGTAATATCTGTTGTTGCTGTAAGAGTGTGTAATTTCAAAAATCTGAAATCTCATA 5580
Qy 5581 AAGTTAAATTTTGAAAA 5597
Db 5581 AAGTTAAATTTTGAAAA 5597

RESULT 4

US-09-986-552-4
; Sequence 4, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 4									
; LENGTH: 5597									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-09-986-552-4									
Query Match									
Best Local Similarity 100.0%; Score 5597; DB 4; Length 5597;									
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	CGGAGCCGAGCGGCGCTCCGTCGGCGAGCTGCAATGAGCGGCGCCCGGAGGCTGTGACC	60						
DB	1	CGGAGCCGAGCGGCGCTCCGTCGGCGAGCTGCAATGAGCGGCGCCCGGAGGCTGTGACC	60						
QY	61	TGCGGCGGCGGCGGCGGCGGCGGCGCTGAAATGGCGGCTCGCTGAGCGGCGGCGGCGGC	120						
DB	61	TGCGGCGGCGGCGGCGGCGGCGGCGCTGAAATGGCGGCTCGCTGAGCGGCGGCGGCGGC	120						
QY	121	GCGCGCTCAGGCTCTCTCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT	180						
DB	121	GCGCGCTCAGGCTCTCTCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT	180						
QY	181	TGCGAGACAAACCTATACCTCCGCTGCCAGGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT	240						
DB	181	TGCGAGACAAACCTATACCTCCGCTGCCAGGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT	240						
QY	241	GCGTGGTGTACCATCGTCTCGGCTTCCAGTTCGAGAGGCTGCTCGAATGGAGCC	300						
DB	241	GCGTGGTGTACCATCGTCTCGGCTTCCAGTTCGAGAGGCTGCTCGAATGGAGCC	300						
QY	301	GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	360						
DB	301	GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	360						
QY	361	AGATCGGCTTGTGTCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT	420						
DB	361	AGATCGGCTTGTGTCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT	420						
QY	421	ATCTTGAAGTACTGAAGAACTACAGCAGGTCAGAGAACTAGAGAGGAGGAGGAGGAGGAGG	480						
DB	421	ATCTTGAAGTACTGAAGAACTACAGCAGGTCAGAGAACTAGAGAGGAGGAGGAGGAGGAGG	480						
QY	481	CAATGAGAGAAATCTTGGGAAAACACACCGGAACTCTAAGAAAGTGAAGAGGAGT	540						
DB	481	CAATGAGAGAAATCTTGGGAAAACACACCGGAACTCTAAGAAAGTGAAGAGGAGT	540						
QY	541	TAGAGTGTGCTTAACACACTGCAATTAAGGTGCAATGCTTGTGCTGCTGCTGCTGCTGCTGCTG	600						
DB	541	TAGAGTGTGCTTAACACACTGCAATTAAGGTGCAATGCTTGTGCTGCTGCTGCTGCTGCTGCTG	600						
QY	601	CAGCAACATCACCTGAGGAGCTGCAATCTCTTTATCTCTTTATCTCTTTATCTCTTTATCTCT	660						
DB	601	CAGCAACATCACCTGAGGAGCTGCAATCTCTTTATCTCTTTATCTCTTTATCTCTTTATCTCT	660						
QY	661	ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTTGTTGTTGTTG	720						
DB	661	ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTTGTTGTTGTTG	720						
QY	721	ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA	780						
DB	721	ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA	780						
QY	781	CAGTATGAGGGGATCTTGTGACACAGATAAGAGTCCCTGAGTTAGTCTTAATGCAAG	840						
DB	781	CAGTATGAGGGGATCTTGTGACACAGATAAGAGTCCCTGAGTTAGTCTTAATGCAAG	840						
QY	841	ATTTGGCTTCTCTGAGTGGATTTCCACCAATTCAGAGGAAACAAATCAACTAAACAA	900						
DB	841	ATTTGGCTTCTCTGAGTGGATTTCCACCAATTCAGAGGAAACAAATCAACTAAACAA	900						
QY	901	AATTGCCAGAAATCTTTTCCTCTAAAGTCAAATGTTGTCAGTTGTTATTCAGAGCCAGTG	960						
DB	901	AATTGCCAGAAATCTTTTCCTCTAAAGTCAAATGTTGTCAGTTGTTATTCAGAGCCAGTG	960						
DB	901	AATTGCCAGAAATCTTTTCCTCTAAAGTCAAATGTTGTCAGTTGTTATTCAGAGCCAGTG	960						
QY	961	TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGCAAACTAAGA	1020						
DB	961	TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGCAAACTAAGA	1020						
QY	1021	AGACATGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1080						
DB	1021	AGACATGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1080						
QY	1081	TCAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGAAGATA	1140						
DB	1081	TCAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGAAGATA	1140						
QY	1141	ACGAGAACTGAGGTACTCATTGCGATCTATCGAGGCAATGACCATGGGTTTGGAAATA	1200						
DB	1141	ACGAGAACTGAGGTACTCATTGCGATCTATCGAGGCAATGACCATGGGTTTGGAAATA	1200						
QY	1201	TTTTCATTGTCAACCAACCGGCGAGATTCATCTCTGGCTGAACCTTTGACAACTCTCGAGTGA	1260						
DB	1201	TTTTCATTGTCAACCAACCGGCGAGATTCATCTCTGGCTGAACCTTTGACAACTCTCGAGTGA	1260						
QY	1261	CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC	1320						
DB	1261	CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC	1320						
QY	1321	CTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTCTCCAGAGTGTATTTACCTTAA	1380						
DB	1321	CTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTCTCCAGAGTGTATTTACCTTAA	1380						
QY	1381	ATGATGATGTCATGTTTGGGAGGATGTCTGCCAGATGATTTTACAGTCACTCCAAAG	1440						
DB	1381	ATGATGATGTCATGTTTGGGAGGATGTCTGCCAGATGATTTTACAGTCACTCCAAAG	1440						
QY	1441	GCAGAGGTTTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1500						
DB	1441	GCAGAGGTTTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1500						
QY	1501	GGATTAAGGATGGCTTATTTGCAAGGCTTGTAAATTAATTCAGCTGCGATGGATGGATGGT	1560						
DB	1501	GGATTAAGGATGGCTTATTTGCAAGGCTTGTAAATTAATTCAGCTGCGATGGATGGATGGT	1560						
QY	1561	GGATTTGCTGGAACAGTGGAGGCTGCTATTTGAGGAGGCTGAGGAGGCTGAGGAGTCTGGA	1620						
DB	1561	GGATTTGCTGGAACAGTGGAGGCTGCTATTTGAGGAGGCTGAGGAGGCTGAGGAGTCTGGA	1620						
QY	1621	GTATTGGAGTTGGACACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA	1680						
DB	1621	GTATTGGAGTTGGACACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA	1680						
QY	1681	ATCAGGATGTCGAAATTTCTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1740						
DB	1681	ATCAGGATGTCGAAATTTCTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1740						
QY	1741	TGCTCTGTTGGTTCATGCTGCGGCTGCGGAGGATCAATTTTCAATGATGATGATGATGATGATG	1800						
DB	1741	TGCTCTGTTGGTTCATGCTGCGGCTGCGGAGGATCAATTTTCAATGATGATGATGATGATGATG	1800						
QY	1801	TGATCTTCTCCAAACAGGATCTCATATATTTTCCAAAGGATGATGATGATGATGATGATGATGATG	1860						
DB	1801	TGATCTTCTCCAAACAGGATCTCATATATTTTCCAAAGGATGATGATGATGATGATGATGATGATG	1860						
QY	1861	TCAGCTTTGCAAGTACGAAAGAGGATGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTG	1920						
DB	1861	TCAGCTTTGCAAGTACGAAAGAGGATGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTG	1920						
QY	1921	TTTCGACATGCTTCTATTGTCGCAACAGTGGGAAACCATCCACTCATATGACAGTGGAA	1980						
DB	1921	TTTCGACATGCTTCTATTGTCGCAACAGTGGGAAACCATCCACTCATATGACAGTGGAA	1980						
QY	1981	TGAATGCCACACATATATTTTAACTCTCAGCTTCAAAATACAAACGATCAAGAGTTCA	2040						
DB	1981	TGAATGCCACACATATATTTTAACTCTCAGCTTCAAAATACAAACGATCAAGAGTTCA	2040						

QY 2041 AAATCAGATACAGTGGAGTGGACACAAGGGGGGACCAAACTGAATTTCTACGGCC 2100
Db 2041 AAATCAGATACAGTGGAGTGGACACAAGGGGGGACCAAACTGAATTTCTACGGCC 2100
QY 2101 AGAAGGGTTACGAATTTAGTTAGTCCCATACACCTTTCCAGAGCGGAATCTTTT 2160
Db 2101 AGAAGGGTTACGAATTTAGTTAGTCCCATACACCTTTCCAGAGCGGAATCTTTT 2160
QY 2161 TTGAGGATATCCCAAGAAAACCGCTTCCGGAAGTTTAAAGACATGATGTTAACTCAA 2220
Db 2161 TTGAGGATATCCCAAGAAAACCGCTTCCGGAAGTTTAAAGACATGATGTTAACTCAA 2220
QY 2221 CAAAGAGAGCCAGAGAGGTGAATAATCCCTCGGTAAATATTTCACTCTCTTCAAAG 2280
Db 2221 CAAAGAGAGCCAGAGAGGTGAATAATCCCTCGGTAAATATTTCACTCTCTTCAAAG 2280
QY 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340
Db 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340
QY 2341 AAGGATACAAATTTGTTCCAGTCCAGTCCCTGAGATCAATTTCTGATGAACCTCAGCATG 2400
Db 2341 AAGGATACAAATTTGTTCCAGTCCAGTCCCTGAGATCAATTTCTGATGAACCTCAGCATG 2400
QY 2401 CTAAATAAAAATCAAGCTATATAACAGATGAACAAATCACAAGTTTGGTGGCTCCAC 2460
Db 2401 CTAAATAAAAATCAAGCTATATAACAGATGAACAAATCACAAGTTTGGTGGCTCCAC 2460
QY 2461 AGGAAAACAGGTTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAAGATTGC 2520
Db 2461 AGGAAAACAGGTTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAAGATTGC 2520
QY 2521 AGAGTTGACCTTTCTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2580
Db 2521 AGAGTTGACCTTTCTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2580
QY 2581 CACCCCTGACCTTGAGACACAGCAAGATTTAGAGTGGAACTCACACCCAAAACCA 2640
Db 2581 CACCCCTGACCTTGAGACACAGCAAGATTTAGAGTGGAACTCACACCCAAAACCA 2640
QY 2641 TAGGGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTCCACTGGAAAAGCCAGA 2700
Db 2641 TAGGGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTCCACTGGAAAAGCCAGA 2700
QY 2701 TCACAAAAGAAAAGAAAATCACAGGAAAGAAAAGACAGTAGAATGGAGAAAATG 2760
Db 2701 TCACAAAAGAAAAGAAAATCACAGGAAAGAAAAGACAGTAGAATGGAGAAAATG 2760
QY 2761 CTGAAAATCACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAG 2820
Db 2761 CTGAAAATCACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAG 2820
QY 2821 ATAGTTACTTGGCTTTTGGCATGGGAGAAAAGTATTTCCAGATCTTCTCGAG 2880
Db 2821 ATAGTTACTTGGCTTTTGGCATGGGAGAAAAGTATTTCCAGATCTTCTCGAG 2880
QY 2881 AAGAAAGTCAATTGAAGACACAAATGGGATCTTCACTGATAGCAAAATATCTGGAGGC 2940
Db 2881 AAGAAAGTCAATTGAAGACACAAATGGGATCTTCACTGATAGCAAAATATCTGGAGGC 2940
QY 2941 AACTAAAAGATACATTTGAGATTCCTCAGATATGTAAATTTCTAAATAGCAAGT 3000
Db 2941 AACTAAAAGATACATTTGAGATTCCTCAGATATGTAAATTTCTAAATAGCAAGT 3000
QY 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCACATGATTCGACGGATTGTA 3060
Db 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCACATGATTCGACGGATTGTA 3060
QY 3061 TCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGACGTCAATTTCAAAAGTGC 3120
Db 3061 TCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGACGTCAATTTCAAAAGTGC 3120

QY 3121 GCCATTCTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180
Db 3121 GCCATTCTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180
QY 3181 AGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
Db 3181 AGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
QY 3241 CTGACAGAGAAATCCGAAACACTGGCTACCGAATTCACGAACTCCGCTTAAAGTTTCAGG 3300
Db 3241 CTGACAGAGAAATCCGAAACACTGGCTACCGAATTCACGAACTCCGCTTAAAGTTTCAGG 3300
QY 3301 ATTTGACAGTCTGGACACATCTAATAATTTGCTCAAAATGCTTCTCTGCTGATATCA 3360
Db 3301 ATTTGACAGTCTGGACACATCTAATAATTTGCTCAAAATGCTTCTCTGCTGATATCA 3360
QY 3361 CGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTACTATGATCCCAACCTGCCACCGG 3420
Db 3361 CGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTACTATGATCCCAACCTGCCACCGG 3420
QY 3421 TCACATAAAGTCTAGTAAACAACTGTAAACCACTAAGTAACTGACAAATCCCAAGCATATA 3480
Db 3421 TCACATAAAGTCTAGTAAACAACTGTAAACCACTAAGTAACTGACAAATCCCAAGCATATA 3480
QY 3481 AGCACAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGA 3540
Db 3481 AGCACAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTTAAATGA 3540
QY 3541 TTGCTACCAAGTTTCTCATGTTGGCGAGTTGGATGACATAGAAAACCTTAGGA 3600
Db 3541 TTGCTACCAAGTTTCTCATGTTGGCGAGTTGGATGACATAGAAAACCTTAGGA 3600
QY 3601 AGTTGCTGCTGGAATGACAACTTGACACAACTCATAAAGATGCTCAGACAGTGAAG 3660
Db 3601 AGTTGCTGCTGGAATGACAACTTGACACAACTCATAAAGATGCTCAGACAGTGAAG 3660
QY 3661 CTGTTCTCAGGACTTCTATGAATCCCATGTTCCCATCTTCCCAATTTGAATCTGCCAA 3720
Db 3661 CTGTTCTCAGGACTTCTATGAATCCCATGTTCCCATCTTCCCAATTTGAATCTGCCAA 3720
QY 3721 GAGAGTACGAAACCGTTTCCCTCATATGATGCTGAGGAGTGAAGGCTTATTCGAG 3780
Db 3721 GAGAGTACGAAACCGTTTCCCTCATATGATGCTGAGGAGTGAAGGCTTATTCGAG 3780
QY 3781 ACAATTTGAAGTTTGGACCCATTGTACTAGCAACATTGATTATGTTTACTATATTTCT 3840
Db 3781 ACAATTTGAAGTTTGGACCCATTGTACTAGCAACATTGATTATGTTTACTATATTTCT 3840
QY 3841 CATTTTGTCTGAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAGGAGATAC 3900
Db 3841 CATTTTGTCTGAGCAGTTAATTTGCACTTAAGCGGAAGATATTTCCAGAGGAGATAC 3900
QY 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAAATCTTCAATTTGAAAACCATCTA 3960
Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAAATCTTCAATTTGAAAACCATCTA 3960
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGATTCCAGAGATGCTTTGTGATGTGAT 4020
Db 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGATTCCAGAGATGCTTTGTGATGTGAT 4020
QY 4021 GCTTAGCAGTTTCCGCGAGAGGAAATATCCAGTACCATCTGTTTGTGCGCATGAA 4080
Db 4021 GCTTAGCAGTTTCCGCGAGAGGAAATATCCAGTACCATCTGTTTGTGCGCATGAA 4080
QY 4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGAAAACCTTGTGTGCGAGCAG 4140
Db 4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGAAAACCTTGTGTGCGAGCAG 4140
QY 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCTCATCTTTTATAAAA 4200
Db 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCTCATCTTTTATAAAA 4200
QY 4201 GGCTCACTGACAAGAGACAGCTGTAAATTTCCCAAGCAATCATTCGAGACTAACTTTAT 4260

4201	Db		GGCTCACTGCAAGAGACAGCTGTTAAATTTCCCAACAGCAATCAATGCGAGACTAACTTTAT	4260
4261	Qy		TAGGAGAAGCCTATGCCAGCTGGGAGTGATTGCTTAAGAGGCTCCAGTCTTTGGATTCCAA	4320
4261	Db		TAGGAGAAGCCTATGCCAGCTGGGAGTGATTGCTTAAGAGGCTCCAGTCTTTGGATTCCAA	4320
4321	Qy		AGCCTTTTGCTAAAGTTTGGACATTTTTTTTTTTTCATTTGCCATTTTTTAAAGTAGTACTA	4380
4321	Db		AGCCTTTTGCTAAAGTTTGGACATTTTTTTTTTTTCATTTGCCATTTTTTAAAGTAGTACTA	4380
4381	Qy		AGTTAACTAGTTATTCTTGTCTCTGAGCTATAACGAATGGAGTCTCTAAACCTATTATTTA	4440
4381	Db		AGTTAACTAGTTATTCTTGTCTCTGAGCTATAACGAATGGAGTCTCTAAACCTATTATTTA	4440
4441	Qy		TAGATGTTATTTAAATAATCAGCAATATCACCTCTTATTGACAATACTTAAATTTATGAG	4500
4441	Db		TAGATGTTATTTAAATAATCAGCAATATCACCTCTTATTGACAATACTTAAATTTATGAG	4500
4501	Qy		TTTTTATTAATTTTAAGACTGTAAATGGTCTTAAACCACTAACTACTGGAAGACTCAATG	4560
4501	Db		TTTTTATTAATTTTAAGACTGTAAATGGTCTTAAACCACTAACTACTGGAAGACTCAATG	4560
4561	Qy		ATTGACATCTCAAAATGCTTTGTAATTATTGACATTCAGCCCTAAGAAATGCTATGATTTCA	4620
4561	Db		ATTGACATCTGAAATGCTTTGTAATTATTGACATTCAGCCCTAAGAAATGCTATGATTTCA	4620
4621	Qy		CTGCGAGGTCTAAATTTCAACAGGCTAGAGTTAGTACTTACCAAGATGTAATATGTTT	4680
4621	Db		CTGCGAGGTCTAAATTTCAACAGGCTAGAGTTAGTACTTACCAAGATGTAATATGTTT	4680
4681	Qy		TGGAATGTAACATATTCAACACAGAGTGCCTCATTTTAGAAATGAGTGTGCTGATGGCA	4740
4681	Db		TGGAATGTAACATATTCAACACAGAGTGCCTCATTTTAGAAATGAGTGTGCTGATGGCA	4740
4741	Qy		CTGSCAATTAACAGTGGTGTCTTTGTTAAATACTACTCATTTGGTATATTCAGTAGTACTCTCT	4800
4741	Db		CTGSCAATTAACAGTGGTGTCTTTGTTAAATACTACTCATTTGGTATATTCAGTAGTACTCTCT	4800
4801	Qy		CTCAGTTGGTTTTTGATAGAACAGAGCCACAGCAAACTTTCTTTGTTAAAGGCTGGTTAGT	4860
4801	Db		CTCAGTTGGTTTTTGATAGAACAGAGCCACAGCAAACTTTCTTTGTTAAAGGCTGGTTAGT	4860
4861	Qy		AAATATTGACGGCCACTGTGTCTTTGTGCATACATTTCTTTGCTGTTGTTAGTTGT	4920
4861	Db		AAATATTGACGGCCACTGTGTCTTTGTGCATACATTTCTTTGCTGTTGTTAGTTGT	4920
4921	Qy		TTTTTTTTCAACAAACCTCTAAATAATGTAAAAACCAATGTTTAGCTTGACGTGTACAAA	4980
4921	Db		TTTTTTTTCAACAAACCTCTAAATAATGTAAAAACCAATGTTTAGCTTGACGTGTACAAA	4980
4981	Qy		ACTGCCACACAGACAGATGTGACCTCAGGCCATCATTTGCCAATCACTCAGAGAAATTATTT	5040
4981	Db		ACTGCCACACAGACAGATGTGACCTCAGGCCATCATTTGCCAATCACTCAGAGAAATTATTT	5040
5041	Qy		TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTGTGCCAGGCTGGAG	5100
5041	Db		TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTGTGCCAGGCTGGAG	5100
5101	Qy		TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCGGCTCCGGGTTCAAGCAGTTCTGTCTC	5160
5101	Db		TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCGGCTCCGGGTTCAAGCAGTTCTGTCTC	5160
5161	Qy		TCAGCCTTCTGAGTAGCTGGGACTACAGGTGTCATGCCACACACCTCGTAAATTTTGTGA	5220
5161	Db		TCAGCCTTCTGAGTAGCTGGGACTACAGGTGTCATGCCACACACCTCGTAAATTTTGTGA	5220
5221	Qy		TTTTTTAGTAGACGGGGGTTCCACATATTGTCAGGCTTATCTTGAACCTCTGACCTC	5280
5221	Db		TTTTTTAGTAGACGGGGGTTCCACATATTGTCAGGCTTATCTTGAACCTCTGACCTC	5280
5281	Qy		AGGTGATCCACCTGCCTCTGCCTTCCAAAGTGTGAGATTTACAGGCAATAGCCAGGTGCAC	5340

Db	5281	AGTGATGCCACCTGCTCTCCCTCCCAAGTGTGAGATTACAGGCAATAGCCAGTGCAC	5343
Qy	5341	CCAGCCGAGAAATAGTATATTTTATGTATGTTTAAACCTTGGCGTCTAGCCATATTTTATG	5400
Db	5341	CCAGCCGAGAAATAGTATATTTTATGTATGTTTAAACCTTGGCGTCTAGCCATATTTTATG	5400
Qy	5401	TCATAATACAAATGATTTGTGAAGAGCAGATTCCATGAGTAACCTCTGACAGGTATTTTATG	5460
Db	5401	TCATAATACAAATGATTTGTGAAGAGCAGATTCCATGAGTAACCTCTGACAGGTATTTTATG	5460
Qy	5461	ATCATGATCTCAACAATATTCCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA	5520
Db	5461	ATCATGATCTCAACAATATTCCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA	5520
Qy	5521	ATGTAATACTGTTGTTGCTCTGATAGAGTTGTGTAATTTCAAAACTGAAATCTCATAA	5580
Db	5521	ATGTAATACTGTTGTTGCTCTGATAGAGTTGTGTAATTTCAAAACTGAAATCTCATAA	5580
Qy	5581	AAGTTAAATTTTGAAAA 5597	
Db	5581	AAGTTAAATTTTGAAAA 5597	
RESULT 5			
US-09-636-596C-4			
; Sequence 4, Application US/09636596C			
; Patent No. 6770468			
; GENERAL INFORMATION:			
; APPLICANT: CANFIELD, WILLIAM			
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLNcase OF THE LYXOSOMAL TARGETING			
; FILE REFERENCE: 10929-0001-77			
; CURRENT APPLICATION NUMBER: US/09/636,596C			
; CURRENT FILING DATE: 2000-08-10			
; PRIOR APPLICATION NUMBER: 60/153,831			
; PRIOR FILING DATE: 1999-09-14			
; NUMBER OF SEQ ID NOS: 52			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 4			
; LENGTH: 5597			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-636-596C-4			
Query Match 100.0%; Score 5597; DB 4; Length 5597;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	CGAGCGGAGCGGGCGTCCGTGCGCGAGAGCTGCAATGAGCGGGCCCGCGGAGGCTGTGACC	60
Db	1	CGAGCGGAGCGGGCGTCCGTGCGCGAGAGCTGCAATGAGCGGGCCCGCGGAGGCTGTGACC	60
Qy	61	TGCGCGCGGCGGCCCGACCGGGGCCCTGAATGCGCGGCTCGCTGAGCGCGCGCGCGCGCG	120
Db	61	TGCGCGCGGCGGCCCGACCGGGGCCCTGAATGCGCGGCTCGCTGAGCGCGCGCGCGCGCG	120
Qy	121	GGCGGCTCAGGCTCCCTCGGGGGGTGCGGTGAGGGGTGATGCTGTTCAAGCTCC	180
Db	121	GGCGGCTCAGGCTCCCTCGGGGGGTGCGGTGAGGGGTGATGCTGTTCAAGCTCC	180
Qy	181	TGCAGAGACAAACCTATACCTGCCTGCCACAGAGTATGGGCTCTACGCTGCTGCTTCTTG	240
Db	181	TGCAGAGACAAACCTATACCTGCCTGCCACAGAGTATGGGCTCTACGCTGCTGCTTCTTG	240
Qy	241	GCCTCGTTGTCACCATCGCTCCGCCCTCCAGTTCGGAGAGGTGGTCTCTCGAATGAGCC	300
Db	241	GCCTCGTTGTCACCATCGCTCCGCCCTCCAGTTCGGAGAGGTGGTCTCTCGAATGAGCC	300
Qy	301	GAGATCAATACAAATGTTTGTGATTCCTATAGACAAATATTCGTGAAAAGTCTCTTC	360
Db	301	GAGATCAATACAAATGTTTGTGATTCCTATAGACAAATATTCGTGAAAAGTCTCTTC	360
Qy	361	AGAATCGGTTTGTCTGCCCATCGCGATTGACGTTGTTTACACCTCGGTGCAATGGCACAG	420

361	AGAAATCGCGCTTTGTCTCCCATGCCGANTGACGTTGTTTACACTCGTGGTGAATGGCACAG	420
421	ATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGCGAGGAGCAGCAAAAG	480
421	ATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGCGAGGAGCAGCAAAAG	480
481	CAATGACAGAAATCCTTTGGGAAAAACAACAGGNAACCTACTAAGAGAGTGAAGACGAGT	540
481	CAATGACAGAAATCCTTTGGGAAAAACAACAGGNAACCTACTAAGAGAGTGAAGACGAGT	540
541	TAGAGTGTGTGCTAAACACACTGCATTAAAGTGGCCAAATGCTTGTACTTGGACCCAGCCCTGC	600
541	TAGAGTGTGTGCTAAACACACTGCATTAAAGTGGCCAAATGCTTGTACTTGGACCCAGCCCTGC	600
601	CAGCCAAATCACCCCTGAAAGAGCTGTCATCTCTTTATCCCTCTTTTCATTCTGCCAGTG	660
601	CAGCCAAATCACCCCTGAAAGAGCTGTCATCTCTTTATCCCTCTTTTCATTCTGCCAGTG	660
661	ACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGCTCAGTGTGTTGTTTTCG	720
661	ACATTTTCAATGTTGCAAAACCAAAAACCCCTTCTACCAATGCTCAGTGTGTTGTTTTCG	720
721	ACAGTACTAAGGATGTTGAAGATGCCCACTCTGCACTGCTTAAAGGAAATPAGCAGACAGA	780
721	ACAGTACTAAGGATGTTGAAGATGCCCACTCTGCACTGCTTAAAGGAAATPAGCAGACAGA	780
781	CAGTATCGAGGGGGTACTTGACACACATAAAGAGTCCCTGGATAGTCTACTAATCCAAG	840
781	CAGTATCGAGGGGGTACTTGACACACATAAAGAGTCCCTGGATAGTCTACTAATCCAAG	840
841	ATTTGGCTTTCCCTGAGTGGATTTCCACCAACATTCAAGGAAAAAAATCAACTAAAAACAA	900
841	ATTTGGCTTTCCCTGAGTGGATTTCCACCAACATTCAAGGAAAAAAATCAACTAAAAACAA	900
901	AATTGCCAGAAAATCTTTTCCCTTAAAGTCAAACTGTGTTGCAAGTTGTATTCAGAGCCAGTG	960
901	AATTGCCAGAAAATCTTTTCCCTTAAAGTCAAACTGTGTTGCAAGTTGTATTCAGAGCCAGTG	960
961	TAGCGCTCTTAAAACTGAATAACCCCAAGANTTTCAAGAAATTTGAATAAGCAAACTAAGA	1020
961	TAGCGCTCTTAAAACTGAATAACCCCAAGANTTTCAAGAAATTTGAATAAGCAAACTAAGA	1020
1021	AGAACTAGCACTGATGATGGAAGAAAGACTGACCATPAAGTCTCTGCATATTTATTATGGATC	1080
1021	AGAACTAGCACTGATGATGGAAGAAAGACTGACCATPAAGTCTCTGCATATTTATTATGGATC	1080
1081	TGACGGCCATCAGCCAGCTCTAAGCAGAGATGAAGACATCTCTGCCAGTCGTTTGAAGATA	1140
1081	TGACGGCCATCAGCCAGCTCTAAGCAGAGATGAAGACATCTCTGCCAGTCGTTTGAAGATA	1140
1141	ACGAAGAACTGAGGTACTCATTTGGATCTATTCGAGAGGCATGCACCATGGTTTCGGNAAT	1200
1141	ACGAAGAACTGAGGTACTCATTTGGATCTATTCGAGAGGCATGCACCATGGTTTCGGNAAT	1200
1201	TTTTTANTTGTCAACAAAGGGCAGATTCCATCTCGGCTGAACCTTGAACAATCTCTCGAGTGA	1260
1201	TTTTTANTTGTCAACAAAGGGCAGATTCCATCTCGGCTGAACCTTGAACAATCTCTCGAGTGA	1260
1261	CAATAGTAAACACACAGAGTGTTTTTCGAAATTTGAGCCACTTGCCCTACCTTTAGTTCAC	1320
1261	CAATAGTAAACACACAGAGTGTTTTTCGAAATTTGAGCCACTTGCCCTACCTTTAGTTCAC	1320
1321	CTGCTATTGAAAGTCAATTATCGCATTCGAAGGGCTGTCCAGAGATTTATTATCTCTAA	1380
1321	CTGCTATTGAAAGTCAATTATCGCATTCGAAGGGCTGTCCAGAGATTTATTATCTCTAA	1380
1381	ATGATGATGTCACTGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCCAAAG	1440
1381	ATGATGATGTCACTGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCCAAAG	1440
1441	GCCAGAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCCGAGTTCTCT	1500
1441	GCCAGAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCCGAGTTCTCT	1500

QY	1501	GGATTAAAGATGGCTATTGTGCAAGAGCCTTGTAAATAATTGAGCCTCGGATTTGGATGGTG	1560
DB	1501	GGATTAAAGATGGCTATTGTGCAAGAGCCTTGTAAATAATTGAGCCTCGGATTTGGATGGTG	1560
QY	1561	GGGATTGCTCTGGAAAAAGTGGAGGAGTCGCTATATTGCAAGAGGTGGAGGTACTGGGA	1620
DB	1561	GGGATTGCTCTGGAAAAAGTGGAGGAGTCGCTATATTGCAAGAGGTGGAGGTACTGGGA	1620
QY	1621	GTATTGAGATTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTGA	1680
DB	1621	GTATTGAGATTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTGA	1680
QY	1681	ATCAGGAGATGTCGAAATTCCTGGCTCGCTGATAAAGTTCGTGACCAAGCATGCAATGTCT	1740
DB	1681	ATCAGGAGATGTCGAAATTCCTGGCTCGCTGATAAAGTTCGTGACCAAGCATGCAATGTCT	1740
QY	1741	TGTCCTGTGGGTTTGATGCTGGCGACTGTGGGCGAGATCAATTTTCATGAATTTGTATAAAG	1800
DB	1741	TGTCCTGTGGGTTTGATGCTGGCGACTGTGGGCGAGATCAATTTTCATGAATTTGTATAAAG	1800
QY	1801	TGATCCTTCTCCAAACACAGACTCACATATATTATCCAAAAGGTGAATGCCTGCTTATT	1860
DB	1801	TGATCCTTCTCCAAACACAGACTCACATATATTATCCAAAAGGTGAATGCCTGCTTATT	1860
QY	1861	TCAGCTTTGCAGAGTAGCCMAAGAGGAGTTGAGGTGCCTATAGTGACAAATCCANTAA	1920
DB	1861	TCAGCTTTGCAGAGTAGCCMAAGAGGAGTTGAGGTGCCTATAGTGACAAATCCANTAA	1920
QY	1921	TTGCACATGCTTCTATTGCCAAACAAGTGGAAAAACCATCCACTCATATATGCACAGTGGAA	1980
DB	1921	TTGCACATGCTTCTATTGCCAAACAAGTGGAAAAACCATCCACTCATATATGCACAGTGGAA	1980
QY	1981	TGAATGCCACCAACAATCAATTTAATCTCAGCTTCAAAATACAAACGATGAGAGTTCA	2040
DB	1981	TGAATGCCACCAACAATCAATTTAATCTCAGCTTCAAAATACAAACGATGAGAGTTCA	2040
QY	2041	AAATGCAGATAAACAGTGGAGGTGCACAACAGGGAGGACCAAAACTCAAATCTACGGCCC	2100
DB	2041	AAATGCAGATAAACAGTGGAGGTGCACAACAGGGAGGACCAAAACTCAAATCTACGGCCC	2100
QY	2101	AGAAGGGTTACGAAAAATTTAGTTAGTCCCATACACTTCTTCAGAGGCGGAATCCTTT	2160
DB	2101	AGAAGGGTTACGAAAAATTTAGTTAGTCCCATACACTTCTTCAGAGGCGGAATCCTTT	2160
QY	2161	TTGAGGATATTTCCCAAGAAAAACCGTTCCTCCGAAGTTTAAAGACATGATGTTAACTCAA	2220
DB	2161	TTGAGGATATTTCCCAAGAAAAACCGTTCCTCCGAAGTTTAAAGACATGATGTTAACTCAA	2220
QY	2221	CAAGAGAGCCAGGAAGAGGTGAAATTCCTGTTAAATATTCTACTCCTTCCAAAAG	2280
DB	2221	CAAGAGAGCCAGGAAGAGGTGAAATTCCTGTTAAATATTCTACTCCTTCCAAAAG	2280
QY	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTGTGAACTGGAACATGGAGACATCACTTTGA	2340
DB	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTGTGAACTGGAACATGGAGACATCACTTTGA	2340
QY	2341	AAGATACAATTTGTCCAAGTCAGCCTTGCTGAGATCAATTTCTGATGAATCAGAGATG	2400
DB	2341	AAGATACAATTTGTCCAAGTCAGCCTTGCTGAGATCAATTTCTGATGAATCAGAGATG	2400
QY	2401	CTAAAAATAAAAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCAC	2460
DB	2401	CTAAAAATAAAAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCAC	2460
QY	2461	AGGAAACAGGTTTCATAAAGCATCTTGCAACACGCTTAGGAGTCTCTGAAAGATTGC	2520
DB	2461	AGGAAACAGGTTTCATAAAGCATCTTGCAACACGCTTAGGAGTCTCTGAAAGATTGC	2520
QY	2521	AGAGTTGACTTTTCTCGAGTAGTGTAAAAAGTGAATGTGTCATGACCGGTCAGAAATC	2580
DB	2521	AGAGTTGACTTTTCTCGAGTAGTGTAAAAAGTGAATGTGTCATGACCGGTCAGAAATC	2580

QY	2581	CACCCCTGGAGCTTGGAGACCAACAGATTTAGAGTGGAACTCAACACCAAAAAACCA	2640	Db	3661	CTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACATGCCAA	3720
Db	2581	CACCCCTGGAGCTTGGAGACCAACAGATTTAGAGTGGAACTCAACACCAAAAAACCA	2640	QY	3721	GAGAGTATCGAAAACGTTTCTTCATATGCATGAGCTGCAGGAATGAGAGGCTTATCGAG	3780
QY	2641	TAGGCGGAAATGTGACAAAGAAAGACCCCATCTCTGATTGTTCCACTGCGAAGCCAGA	2700	Db	3721	GAGAGTATCGAAAACGTTTCTTCATATGCATGAGCTGCAGGAATGAGAGGCTTATCGAG	3780
Db	2641	TAGGCGGAAATGTGACAAAGAAAGACCCCATCTCTGATTGTTCCACTGCGAAGCCAGA	2700	QY	3781	ACAAATGGAAGTTTGGACCCCATTTGTGTACTAGCAACATTTGATTAAGTTTACTATTTCT	3840
QY	2701	TGACAAAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGTAGAATGAGGAAATG	2760	Db	3781	ACAAATGGAAGTTTGGACCCCATTTGTGTACTAGCAACATTTGATTAAGTTTACTATTTCT	3840
Db	2701	TGACAAAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGTAGAATGAGGAAATG	2760	QY	3841	CATTTTCTCAGCAGATTAATTTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
QY	2761	CTGAAATCACAATAGGCGCTTACTGAGTGTACTTGGNAGAAAGCTGCAGCAATACACAG	2820	Db	3841	CATTTTCTCAGCAGATTAATTTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
Db	2761	CTGAAATCACAATAGGCGCTTACTGAGTGTACTTGGNAGAAAGCTGCAGCAATACACAG	2820	QY	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAACCATCTA	3960
QY	2821	ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTCGACG	2880	Db	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAACCATCTA	3960
Db	2821	ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTCGACG	2880	QY	3961	CCTCAGCATTTACTCAGCAGATTTTAAACTCAGCTTTCACAGAGATGCTTTGTGTGATGAT	4020
QY	2881	AAGAGAGTCAATGAGACACAATTTGGCAATCTTCACTGATAGCAAAAAATCTGGGAGGC	2940	Db	3961	CCTCAGCATTTACTCAGCAGATTTTAAACTCAGCTTTCACAGAGATGCTTTGTGTGATGAT	4020
Db	2881	AAGAGAGTCAATGAGACACAATTTGGCAATCTTCACTGATAGCAAAAAATCTGGGAGGC	2940	QY	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
QY	2941	AACATAAAGATACATTTCCAGATTCCTTCAGATATGTAAATAAAATTTCTAAATAGCAAGT	3000	Db	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
Db	2941	AACATAAAGATACATTTCCAGATTCCTTCAGATATGTAAATAAAATTTCTAAATAGCAAGT	3000	QY	4081	FATAGCCCACTGACTAGGAATTTTAAACCAACCACCTGAAAACTTTGTGTGTCGAGGAC	4140
QY	3001	TTGGATTCACATTCGCGGAAATTCCTGCTCACAATGCTCAGATGATGACCGGATGTTA	3060	Db	4081	FATAGCCCACTGACTAGGAATTTTAAACCAACCACCTGAAAACTTTGTGTGTCGAGGAC	4140
Db	3001	TTGGATTCACATTCGCGGAAATTCCTGCTCACAATGCTCAGATGATGACCGGATGTTA	3060	QY	4141	TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTCTTTTATAAAA	4200
QY	3061	TGCAGAACTCAAGATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTTCAAAAATGTC	3120	Db	4141	TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCTCTTTTATAAAA	4200
Db	3061	TGCAGAACTCAAGATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTTCAAAAATGTC	3120	QY	4201	GGCTCAGTGAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGAGTACTTTAT	4260
QY	3121	GCCATTCAGAGATATGAGTTTGCCTTCTCTTATTTTATTTATPCTCATGAGTGCAGTGC	3180	Db	4201	GGCTCAGTGAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGAGTACTTTAT	4260
Db	3121	GCCATTCAGAGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC	3180	QY	4261	TAGGAGAGCCCTATGCCAGCTGGGAGTCAATGCTAAAGAGGCTCCAGTCTTTTGCATTC	4320
QY	3181	AGCCACTGAATATATCTCAAGTCTTGTGATGAGTGTATACAGATCAATCTGTTCTGT	3240	Db	4261	TAGGAGAGCCCTATGCCAGCTGGGAGTCAATGCTAAAGAGGCTCCAGTCTTTTGCATTC	4320
Db	3181	AGCCACTGAATATATCTCAAGTCTTGTGATGAGTGTATACAGATCAATCTGTTCTGT	3240	QY	4321	AGCCCTTTTCTAAAGTTTGCACCTTTTTTTTTCATTTCCCATTTTTPAAGTAGTACTA	4380
QY	3241	CTGACAGAGAAATCCGAACACTGGCTACCAAGATTCACGAACCTGCCGTTAAGTTTGCAGG	3300	Db	4321	AGCCCTTTTCTAAAGTTTGCACCTTTTTTTTTCATTTCCCATTTTTPAAGTAGTACTA	4380
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACCAAGATTCACGAACCTGCCGTTAAGTTTGCAGG	3300	QY	4381	AGTTAACTAGTTTATCTTCTGAGTATACGAATTTGGGATGTCCTAAACCTATTTT	4440
QY	3301	ATTTGACAGGCTCGGAACACATGCTATATAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360	Db	4381	AGTTAACTAGTTTATCTTCTGAGTATACGAATTTGGGATGTCCTAAACCTATTTT	4440
Db	3301	ATTTGACAGGCTCGGAACACATGCTATATAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360	QY	4441	TAGATGTTTATTAATAATGAGCAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
QY	3361	CGCAGTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG	3420	Db	4441	TAGATGTTTATTAATAATGAGCAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
Db	3361	CGCAGTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG	3420	QY	4501	TTTTTATTAATTTTAAAGCTGTAATTTGCTTAAACCACTAACCTACTGAGAGCTCAATG	4560
QY	3421	TCACATAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATA	3480	Db	4501	TTTTTATTAATTTTAAAGCTGTAATTTGCTTAAACCACTAACCTACTGAGAGCTCAATG	4560
Db	3421	TCACATAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATA	3480	QY	4561	ATTGACATCTGAAATGCTTTTGAATTTATTTGATCTTCAGCCCTCCTAAGATGCTATGATTC	4620
QY	3481	AGGACAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540	Db	4561	ATTGACATCTGAAATGCTTTTGAATTTATTTGATCTTCAGCCCTCCTAAGATGCTATGATTC	4620
Db	3481	AGGACAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540	QY	4621	CCTGAGGCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT	4680
QY	3541	TTCTGTACCAACGTTTCTCATGTGTTGGCCAGTTTGGATGACATAAGAAAAAACCCCTAGGA	3600	Db	4621	CCTGAGGCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT	4680
Db	3541	TTCTGTACCAACGTTTCTCATGTGTTGGCCAGTTTGGATGACATAAGAAAAAACCCCTAGGA	3600	QY	4681	TGGAATGTACATATTCACACAGAGTGCCTCATTTTGAAGAAATGAGTAGTCTGATGGCA	4740
QY	3601	AGTTTGTTCCTCAATGACAACTTGCACCAATCATTAAGATGCTTCAGACAGTGAAGG	3660	Db	4681	TGGAATGTACATATTCACACAGAGTGCCTCATTTTGAAGAAATGAGTAGTCTGATGGCA	4740
Db	3601	AGTTTGTTCCTCAATGACAACTTGCACCAATCATTAAGATGCTTCAGACAGTGAAGG	3660	QY	4741	CTGGCACATTCAGTGGTCTTGTTTTAATACTCATTTGTTATATTCAGTAGTACTTCTCT	4800
QY	3661	CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA	3720				

Db 4741 CTGGCACATTACAGTGGTGTCTGTTTAACTACTCATTTGGTATATTCACGTAGCTATCTCT 4800
Qy 4801 CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTAGT 4860
Db 4801 CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTAGT 4860
Qy 4861 AAATTAATTGACGGCCACCTGTGCTTTGTCATACATTTCTTCTGCTGTGTTAGTTTGT 4920
Db 4861 AAATTAATTGACGGCCACCTGTGCTTTGTCATACATTTCTTCTGCTGTGTTAGTTTGT 4920
Qy 4921 TTTTTTCAAAACACCCCTCTAAAAATGTAATAAATGTAATAAATGTAATAAATGTAATAA 4980
Db 4921 TTTTTTCAAAACACCCCTCTAAAAATGTAATAAATGTAATAAATGTAATAAATGTAATAA 4980
Qy 4981 ACTGCCACCCAGCCAGATGTGACCTCAGGCCATCATTTGCCATCACTGAGAAATATTT 5040
Db 4981 ACTGCCACCCAGCCAGATGTGACCTCAGGCCATCATTTGCCATCACTGAGAAATATTT 5040
Qy 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100
Qy 5101 TGCAGTGGCGGAATCTCAGCTCACTGCAACCTCCGCGCTCCGCGGTTCAAGCAGTTCTGTC 5160
Db 5101 TGCAGTGGCGGAATCTCAGCTCACTGCAACCTCCGCGCTCCGCGGTTCAAGCAGTTCTGTC 5160
Qy 5161 TCAGCTTTCTGAGTAGCTGGGACTACAGGTGTCATGCCACACACCCCTGCTAAATTTTGTGTA 5220
Db 5161 TCAGCTTTCTGAGTAGCTGGGACTACAGGTGTCATGCCACACACCCCTGCTAAATTTTGTGTA 5220
Qy 5221 TTTTTTAGTAGAGACGGGGGTTCCACCATATTTGGTCAGGCTTTATCTTTGAACCTCTGACCTC 5280
Db 5221 TTTTTTAGTAGAGACGGGGGTTCCACCATATTTGGTCAGGCTTTATCTTTGAACCTCTGACCTC 5280
Qy 5281 AGGTGATCCACCTGCTGCTCCCAAGTCTGAGATTACAGGCATACGCCAGTGCAC 5340
Db 5281 AGGTGATCCACCTGCTGCTCCCAAGTCTGAGATTACAGGCATACGCCAGTGCAC 5340
Qy 5341 CCAGCCGAGAAATAGTATTTTATGTTGTTAAACCTTGGGCTGCTAGCCATATTTTATG 5400
Db 5341 CCAGCCGAGAAATAGTATTTTATGTTGTTAAACCTTGGGCTGCTAGCCATATTTTATG 5400
Qy 5401 TCATTAATCAATGATTTGTTGAGAGCAGATTCATGATGTAATCTGACAGTATTTTAG 5460
Db 5401 TCATTAATCAATGATTTGTTGAGAGCAGATTTCCATGATGTAATCTGACAGTATTTTAG 5460
Qy 5461 ATCATGATCTCAACAATATCTCCCAATGGCATACATCTTTTGTACAAAAGACTTGAA 5520
Db 5461 ATCATGATCTCAACAATATCTCCCAATGGCATACATCTTTTGTACAAAAGACTTGAA 5520
Qy 5521 ATGTAATACTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5580
Db 5521 ATGTAATACTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5580
Qy 5581 AAGTTAAATTTTGAATA 5597
Db 5581 AAGTTAAATTTTGAATA 5597

RESULT 6

US-09-635-872A-20
; Sequence 20, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613USO
; CURRENT APPLICATION NUMBER: US/09/635, 872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-635-872A-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 165 ATGCTGTTCAAGCTCTGCGAGAGACAACTTACCTGCTGTCACAGGTATGGGCTC 224
Db 7 ATGGGGTTCAAGCTCTGCGAGAGACAACTTACCTGCTGTCACAGGTATGGGCTC 66

Qy 225 TAGCTGTGCTTCTTGGGCGTCTGTTGTCACATCTGCTCCGCTTCCAGTTCGAGAGG 284
Db 67 TAGCTGTGCTTCTTGGGCGTCTGTTGTCACATCTGCTCCGCTTCCAGTTCGAGAGG 126

Qy 285 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 344
Db 127 GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTT 186

Qy 345 GCTGGAAAGTCTTTTCCAGAAATCGGCTTGTCTGCCCATGCCGATTCGAGTTGTTTAC 404
Db 187 GCTGGAAAGTCTTTTCCAGAAATCGGCTTGTCTGCCCATGCCGATTCGAGTTGTTTAC 246

Qy 405 TGGGTGAATGGCACAGATCTTGAATCTACTGAAGAACTACAGCAGGTACAGAAACAGATG 464
Db 247 TGGGTGAATGGCACAGATCTTGAATCTACTGAAGAACTACAGCAGGTACAGAAACAGATG 306

Qy 465 GAGGAGAGCAGAAAGCAATGAGAGAAATCCCTTGGGAAACACAAACCGAACTCTAAG 524
Db 307 GAGGAGAGCAGAAAGCAATGAGAGAAATCCCTTGGGAAACACAAACCGAACTCTAAG 366

Qy 525 AAGAGTGAGAGCAGTTAGAGTGTGTTGCTAAACACATGCTGATTAAGTGGCAATGCTTGA 584
Db 367 AAGAGTGAGAGCAGTTAGAGTGTGTTGCTAAACACATGCTGATTAAGTGGCAATGCTTGA 426

Qy 585 CTGAGCCCGCCCTGCGCAGCAACATCACCTCTGAAGGACGTGCCATCTCTTATCTCTT 644
Db 427 CTGAGCCCGCCCTGCGCAGCAACATCACCTCTGAAGGACGTGCCATCTCTTATCTCTT 486

Qy 645 TTTCTATCTGCCAGTACATTTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGTC 704
Db 487 TTTCTATCTGCCAGTACATTTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGTC 546

Qy 705 TCAGTTGTTGTTTTCACAGTACTAAGAGTGTGTAAGATGCCCATCTCTGAGCTGCTTAAA 764
Db 547 TCAGTTGTTGTTTTCACAGTACTAAGAGTGTGTAAGATGCCCATCTCTGAGCTGCTTAAA 606

Qy 765 GGAATAGCAGACAGATGAGGGGGTACTTGACAAACAGATTAAGAAAGTCCCTGGA 824
Db 607 GGAATAGCAGACAGATGAGGGGGTACTTGACAAACAGATTAAGAAAGTCCCTGGA 666

Qy 825 TTAGTGTAAATGCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACA 884
Db 667 TTAGTGTAAATGCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACA 726

Qy 885 AATCAACTAAAAACAAAATGCGAGAAATCTTCTCTTAAAGTCAAACTGTTGCAAGTTG 944
Db 727 AATCAACTAAAAACAAAATGCGAGAAATCTTCTCTTAAAGTCAAACTGTTGCAAGTTG 786

Qy 945 TATTCAGAGCCAGTGTAGGCTTCTTAAACTGAATACCCCAAGGATTTTCAAGAAATG 1004
Db 787 TATTCAGAGCCAGTGTAGGCTTCTTAAACTGAATACCCCAAGGATTTTCAAGAAATG 846

Qy 1005 AATAAGCAAACTAAGAGAAACATGACCATTTGATGTAAGAAAGAACTGACCAATAGTCTGCA 1064
Db 847 AATAAGCAAACTAAGAGAAACATGACCATTTGATGTAAGAAAGAACTGACCAATAGTCTGCA 906

Qy 1065 TATTTATATGAGATCTGAGCGCCATCAGCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124
Db 907 TATTTATATGAGATCTGAGCGCCATCAGCAGTCTAAGCAGGATGAAGACATCTCTGCC 966

QY 1125 AGTCGTTTTGAAGATAACGAAAGAACTGAGGTACTCATTTGGATCTATCGAGAGCATGCA 1184
DB 967 AGTCGTTTTGAAGATAACGAAAGAACTGAGGTACTCATTTGGATCTATCGAGAGCATGCA 1026
QY 1185 CCATGGGTTCCGAATATTTTCATTTGTACCAACGGGAGATTCCATCTCGCTGAACTTT 1244
DB 1027 CCATGGGTTCCGAATATTTTCATTTGTACCAACGGGAGATTCCATCTCGCTGAACTTT 1086
QY 1245 GCAATCTCGAGTGACAAATAGTAAACACACAGAGATGTTTTTTCGAATTTGAGCCATTG 1304
DB 1087 GCAATCTCGAGTGACAAATAGTAAACACACAGAGATGTTTTTTCGAATTTGAGCCATTG 1146
QY 1305 CCTACCTTTAGTTCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAG 1364
DB 1147 CCTACCTTTAGTTCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAG 1206
QY 1365 AAGTTTATTTACCTAAATGATGTGCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1424
DB 1207 AAGTTTATTTACCTAAATGATGTGCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1266
QY 1425 TACAGTCACTCCAAAGGCCAGAGGTTTATTGTGACATGGCTGTGCCAACTGTCCGAG 1484
DB 1267 TACAGTCACTCCAAAGGCCAGAGGTTTATTGTGACATGGCTGTGCCAACTGTCCGAG 1326
QY 1485 GGCTGCCAGGTTCTCGATTAAGGATGGCTATTGTGACAAAGGCTTGTATAATTGAGCC 1544
DB 1327 GGCTGCCAGGTTCTCGATTAAGGATGGCTATTGTGACAAAGGCTTGTATAATTGAGCC 1386
QY 1545 TCGATTTGGGATGTTGGGATGTCCTCGAAACAGTGGAGGAGTCCCTATATTGACGA 1604
DB 1387 TCGATTTGGGATGTTGGGATGTCCTCGAAACAGTGGAGGAGTCCCTATATTGACGA 1446
QY 1605 GGTGAGGATCTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC 1664
DB 1447 GGTGAGGATCTGGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGANTAAAC 1506
QY 1665 AGTGCTCTTAAGTAACGAGGATGTCGGAATTTCTGGCTGCTGTATGATGATGTTCTGAC 1724
DB 1507 AGTGCTCTTAAGTAACGAGGATGTCGGAATTTCTGGCTGCTGTATGATGATGTTCTGAC 1566
QY 1725 CAAGCATCAATGCTTCTGCTGGGTTTGCATGCTGCGACTGTGGGCAAGATCATTTT 1784
DB 1567 CAAGCATCAATGCTTCTGCTGGGTTTGCATGCTGCGACTGTGGGCAAGATCATTTT 1626
QY 1785 CATGAATTTGATAAGTGAATCTTCTCCAAACAGACTCACTATATTATTTCCAAAGGT 1844
DB 1627 CATGAATTTGATAAGTGAATCTTCTCCAAACAGACTCACTATATTATTTCCAAAGGT 1686
QY 1845 GAATGCTGCTTATTTTTCAGGTTTGCAGAGTGCAGAAAGAGGAGTTGAAGTGCTAT 1904
DB 1687 GAATGCTGCTTATTTTTCAGGTTTGCAGAGTGCAGAAAGAGGAGTTGAAGTGCTAT 1746
QY 1905 AGTGACATCAATAATTTCGATGCTTCTATTTCCCAACAGTGGAAACCATCCACCTC 1964
DB 1747 AGTGACATCAATAATTTCGATGCTTCTATTTCCCAACAGTGGAAACCATCCACCTC 1806
QY 1965 ATATGCACTGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2024
DB 1807 ATATGCACTGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1866
QY 2025 AACGATGAAGGTTTCAAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2084
DB 1867 AACGATGAAGGTTTCAAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1926
QY 2085 CTGAATTTACGCGCCAGAGGGTTACGAAATTTAGTTAGTCCATAAATCTTTTCCA 2144
DB 1927 CTGAATTTACGCGCCAGAGGGTTACGAAATTTAGTTAGTCCATAAATCTTTTCCA 1986
QY 2145 GAGCGGAAATCTTTTGGAGATTTCCCAAGAAAGAGGCTTCCGGAAGTTTAAGAGA 2204
DB 1987 GAGCGGAAATCTTTTGGAGATTTCCCAAGAAAGAGGCTTCCGGAAGTTTAAGAGA 2046

QY 2205 CATGATGTTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCTGTAATATT 2264
DB 2047 CATGATGTTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCTGTAATATT 2106
QY 2265 TCACCTCTTTCCAAAAGAGCGCCAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACAT 2324
DB 2107 TCACCTCTTTCCAAAAGAGCGCCAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACAT 2166
QY 2325 GGAGACATCACTTTGAAAGGATCAATTTGTCCAAAGTCAAGTCTGCTGAGATCAATTTCTG 2384
DB 2167 GGAGACATCACTTTGAAAGGATCAATTTGTCCAAAGTCAAGTCTGCTGAGATCAATTTCTG 2226
QY 2385 ATGAACTCACACATCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2444
DB 2227 ATGAACTCACACATCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2286
QY 2445 AGTTTGGTGGTCCACAGGAAACAGGTTTCAATAAAGATCTTTGCCAAAACAGCTTAGGA 2504
DB 2287 AGTTTGGTGGTCCACAGGAAACAGGTTTCAATAAAGATCTTTGCCAAAACAGCTTAGGA 2346
QY 2505 GTGCTCGAAAGATTCAGAGGTTGACTTTTCTGTCAGTGAAGTGAATAAGTGAATGCTCAT 2564
DB 2347 GTGCTCGAAAGATTCAGAGGTTGACTTTTCTGTCAGTGAAGTGAATAAGTGAATGCTCAT 2406
QY 2565 GACCAAGGTCAGAAATCCACCCCTGCACTTGGAGACCACAGCAAGATTTAGAGTGGAAACT 2624
DB 2407 GACCAAGGTCAGAAATCCACCCCTGCACTTGGAGACCACAGCAAGATTTAGAGTGGAAACT 2466
QY 2625 CACACCCAAAACCATAGGCGGAATGTGACAAAGAAAGCCCACTCTCTGATTTGTT 2684
DB 2467 CACACCCAAAACCATAGGCGGAATGTGACAAAGAAAGCCCACTCTCTGATTTGTT 2526
QY 2685 CCACTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGT 2744
DB 2527 CCACTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGT 2586
QY 2745 AGAATGGAGAAATGCTGAAATTCACATAGCGTTACTGAAGTGTACTTTGGAGAAAG 2804
DB 2587 AGAATGGAGAAATGCTGAAATTCACATAGCGTTACTGAAGTGTACTTTGGAGAAAG 2646
QY 2805 CTGACGACATTCACAGATAGTTACTTTGGCTTTTTCATCGGAGAAAGAAAGTATTTT 2864
DB 2647 CTGACGACATTCACAGATAGTTACTTTGGCTTTTTCATCGGAGAAAGAAAGTATTTT 2706
QY 2865 CAAATCTTCTCGAGAAAGAGATGATTTGAAGACAAATTTGGCATCTTCACTGATAGC 2924
DB 2707 CTAGATCTTCTCGAGAAAGAGATGATTTGAAGACAAATTTGGCATCTTCACTGATAGC 2766
QY 2925 AAAAAATCTGGGAGGCACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAA 2984
DB 2767 AAAAAATCTGGGAGGCACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAA 2826
QY 2985 ATTTAAATAGCAAGTTTGGATTCATCTCGGGAAGTCCCTGCTCAATGCCCTCAATG 3044
DB 2827 ATTTAAATAGCAAGTTTGGATTCATCTCGGGAAGTCCCTGCTCAATGCCCTCAATG 2886
QY 3045 ATTTGACCGGATTTGATTCGAGAACTCAAGATATGTTCCCTGAGATTTTGAAGAAAG 3104
DB 2887 ATTTGACCGGATTTGATTCGAGAACTCAAGATATGTTCCCTGAGATTTTGAAGAAAG 2946
QY 3105 TCATTTCAAAAGTGGCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3164
DB 2947 TCATTTCAAAAGTGGCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3006
QY 3165 CTCATGAGTGCAGTCCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224
DB 3007 CTCATGAGTGCAGTCCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066
QY 3225 CAACTCTGCTGCTTCTGTCGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGGAACTG 3284
DB 3067 CAACTCTGCTGCTTCTGTCGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGGAACTG 3126
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGGTTTGCAGGTCGAAACACATGCTAAATAATGCTCAAAAATG 3344

```
Db 3127 CCGTTAAGTTTGACAGATTGACAGTCTGGAACACATGCTAAATAATGCTCAAAAATG 3186
Qy 3345 CTTCTGCTGATATCACGAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3404
Db 3187 CTTCTGCTGATATCACGAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3246
Qy 3405 CCCAACCTGCCACCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAA 3464
Db 3247 CCCAACCTGCCACCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAA 3306
Qy 3465 ATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATCGGAGAGAGAA 3524
Db 3307 ATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATCGGAGAGAGAA 3366
Qy 3525 ATCGCTTTTAAATGATTGTAACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3584
Db 3367 ATCGCTTTTAAATGATTGTAACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3426
Qy 3585 AGAAAAACCCTAGAAGTTGTTTGCCTGAATGACAACTTGACCAANTATAGAT 3644
Db 3427 AGAAAAACCCTAGAAGTTGTTTGCCTGAATGACAACTTGACCAANTATAGAT 3486
Qy 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3704
Db 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3546
Qy 3705 CAATTGAACTCCAGAGAGTATCGAAACCGTTTCCCTTATATGATGAGCTGCGAGAA 3764
Db 3547 CAATTGAACTCCAGAGAGTATCGAAACCGTTTCCCTTATATGATGAGCTGCGAGAA 3606
Qy 3765 TCGAGGGCTTATCGAGACAAATGAAAGTTTGGAGCCATTGTTACTAGCAACATGATT 3824
Db 3607 TCGAGGGCTTATCGAGACAAATGAAAGTTTGGAGCCATTGTTACTAGCAACATGATT 3666
Qy 3825 ATGTTTACTATATCTCATTTTTCCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATT 3884
Db 3667 ATGTTTACTATATCTCATTTTTCCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATT 3726
Qy 3885 CCCAAGAGAGATACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATC 3941
Db 3727 CCCAAGAGAGATACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATC 3783

RESULT 7
US-09-636-077A-20
; Sequence 20, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYXOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-077A-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred No 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 165 ATGCTGTTCAAGCTCTCTGAGAGACAAACCTATACCTGCTGTCCCAAGTATGGGCTC 224
Db 7 ATGGGGTTTCAAGCTCTTTCAGAGACAAACCTATACCTGCTGTCCCAAGTATGGGCTC 66
Qy 225 TACGTGTGCTTCTTGGGGTCTGTTGTCACCATGCTCTCGGCTTCCAGTTCGGAGAGTG 284
```

```
Db 67 TAGCTGTGCTTCTTGGCGCTCGTTGTACCATGCTCTCGCCTTCCAGTTCCGAGAGTG 126
Qy 285 GTTCTGGAATGAGCGGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATT 344
Db 127 GTTCTGGAATGAGCGGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATT 186
Qy 345 GCTGGAAAGTCTTTTTCAGAAATCGGCTTGTCTCCCATCGCGATTGACGTTGTTTACACC 404
Db 187 GCTGGAAAGTCTTTTTCAGAAATCGGCTTGTCTCCCATCGCGATTGACGTTGTTTACACC 246
Qy 405 TGGGTGAATGGCACAGATCTTGAACCTACTGAAGAACTACAGCAGGTACAGAAACAGATG 464
Db 247 TGGGTGAATGGCACAGATCTTGAACCTACTGAAGAACTACAGCAGGTACAGAAACAGATG 306
Qy 465 GAGGAGAGCAGAAAACCAATGAGAGAAATCTTGGGAAAAACAAACGGAACCTACTAAG 524
Db 307 GAGGAGAGCAGAAAACCAATGAGAGAAATCTTGGGAAAAACAAACGGAACCTACTAAG 366
Qy 525 AAGAGTGAGAGCAGTATGAGTGTCTTAAACACACTGCAATTAAGGTGCCAATGTTGA 584
Db 367 AAGAGTGAGAGCAGTATGAGTGTCTTAAACACACTGCAATTAAGGTGCCAATGTTGA 426
Qy 585 CTGAGCCCAAGCTGCGCAGCAACATCACCTCGAAGAGCTGCCATCTCTTTATCCTTCT 644
Db 427 CTGAGCCCAAGCTGCGCAGCAACATCACCTCGAAGAGCTGCCATCTCTTTATCCTTCT 486
Qy 645 TTTCAATCTGCCAGTGACATTTTCAATGTTGAAAAACCAAAACCCCTTCTACCAATGTC 704
Db 487 TTTCAATCTGCCAGTGACATTTTCAATGTTGAAAAACCAAAACCCCTTCTACCAATGTC 546
Qy 705 TCAGTTGTTGTTTTCAGCAGTACTAAGGATGTTGAAGATGTCGCCACTCTGAGCTGCTTAAA 764
Db 547 TCAGTTGTTGTTTTCAGCAGTACTAAGGATGTTGAAGATGTCGCCACTCTGAGCTGCTTAAA 606
Qy 765 GGAATAGCAGACAGACAGTATCGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGA 824
Db 607 GGAATAGCAGACAGACAGTATCGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGA 666
Qy 825 TTAGTCTAATGCAAGATTGGGCTTTCCTGAGTGGATTTCACCAACATTCAGGAAACA 884
Db 667 TTAGTCTAATGCAAGATTGGGCTTTCCTGAGTGGATTTCACCAACATTCAGGAAACA 726
Qy 885 AATCAACTAAAAACAAAATGCCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTTG 944
Db 727 AATCAACTAAAAACAAAATGCCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTTG 786
Qy 945 TATTCAGAGCCAGTGTAGCGCTTCTTAAACTGAATAACCCCAAGATTTCAGAAATTG 1004
Db 787 TATTCAGAGCCAGTGTAGCGCTTCTTAAACTGAATAACCCCAAGATTTCAGAAATTG 846
Qy 1005 AATAAGCAAACTAAGAGAACATGACATTGATGAAAAAGAACTGACCAATAGTCCCTGCA 1064
Db 847 AATAAGCAAACTAAGAGAACATGACATTGATGAAAAAGAACTGACCAATAGTCCCTGCA 906
Qy 1065 TATTTATATGGGATGAGCGCCATCAGCAGTCTTAAGCAGGATGAAGACATCTCTGCC 1124
Db 907 TATTTATATGGGATGAGCGCCATCAGCAGTCTTAAGCAGGATGAAGACATCTCTGCC 966
Qy 1125 AGTCGTTTTCAGACATAACGAAGAACTGAGTACTCATTCGATCTATCGAGAGCATGCA 1184
Db 967 AGTCGTTTTCAGACATAACGAAGAACTGAGTACTCATTCGATCTATTCGAGAGCATGCA 1026
Qy 1185 CCATGGGTTCCGAAATATTTTTCATTGTCAACACGGGAGATTCATCTCGGCTGAAACCTT 1244
Db 1027 CCATGGGTTCCGAAATATTTTTCATTGTCAACACGGGAGATTCATCTCGGCTGAAACCTT 1086
Qy 1245 GACAATCCTCGAGTGACAAATAGTAACACACAGGATGTTTTTGAATTTGAGCACTG 1304
Db 1087 GACAATCCTCGAGTGACAAATAGTAACACACAGGATGTTTTTGAATTTGAGCACTG 1146
Qy 1305 CCTACCTTTAGTTCACTGCTATTGAAAGTCACTTCATCGCATCGAAGGGTGTGCCAG 1364
```

Db 1147 CCTACCTTTAGTTCACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCGCTGCCAG 1206
Qy 1365 RAGTTTATTACCTAAATGATGATGCTCATGTTTGGGAGGATGTCGCCAGATGATTTT 1424
Db 1207 RAGTTTATTACCTAAATGATGATGCTCATGTTTGGGAGGATGTCGCCAGATGATTTT 1266
Qy 1425 TACAGTCTACTCCAAAGGCCAGAGGTTTATTGATCGCTGTGCCAAACTGTGCCGAG 1484
Db 1267 TACAGTCTACTCCAAAGGCCAGAGGTTTATTGATCGCTGTGCCAAACTGTGCCGAG 1326
Qy 1485 GGCTGCCAGGTTCTGCTGATTAAGATGGCTATTGTCACAGGCTTGTATTAATTCAGCC 1544
Db 1327 GGCTGCCAGGTTCTGCTGATTAAGATGGCTATTGTCACAGGCTTGTATTAATTCAGCC 1386
Qy 1545 TCGCATTTGGGATGTTGGGATTTGCTCTGAAACACAGTGGAGGAGTCTGCTATTGTCAGGA 1604
Db 1387 TCGCATTTGGGATGTTGGGATTTGCTCTGAAACACAGTGGAGGAGTCTGCTATTGTCAGGA 1446
Qy 1605 GGTGAGGTACTGGAGTATTGGAGTTGGACACCCCTGGGAGTTTGTGGAGGATTAAC 1664
Db 1447 GGTGAGGTACTGGAGTATTGGAGTTGGACACCCCTGGGAGTTTGTGGAGGATTAAC 1506
Qy 1665 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1724
Db 1507 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1566
Qy 1725 CAAGCATGCAATGCTTCTCTGTGGGTTTGATCTGGCGACTGTGGCGAAGATCATTTT 1784
Db 1567 CAAGCATGCAATGCTTCTCTGTGGGTTTGATCTGGCGACTGTGGCGAAGATCATTTT 1626
Qy 1785 CATGAATTTGATAAGTCACTCTCTCCCAACAGACTCACTATATTATTCCAAAGGT 1844
Db 1627 CATGAATTTGATAAGTCACTCTCTCCCAACAGACTCACTATATTATTCCAAAGGT 1686
Qy 1845 GAATGCCCTGCTTATTTCAGCTTTTCGAGAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1904
Db 1687 GAATGCCCTGCTTATTTCAGCTTTTCGAGAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1746
Qy 1905 AGTGCAATCCAAATTTGACATGCTTCTATTTGCCCAAGTGGAAACCATCCACTC 1964
Db 1747 AGTGCAATCCAAATTTGACATGCTTCTATTTGCCCAAGTGGAAACCATCCACTC 1806
Qy 1965 ATATGACAGTGAATGAATGCCACACACATATTTTAACTCTCAGCTTTCAAATACA 2024
Db 1807 ATATGACAGTGAATGAATGCCACACACATATTTTAACTCTCAGCTTTCAAATACA 1866
Qy 2025 AACGATGAAGGTTCAAATCCAGATTAACAGTGGAGTGGACACAAAGGGAGGCCAAAA 2084
Db 1867 AACGATGAAGGTTCAAATCCAGATTAACAGTGGAGTGGACACAAAGGGAGGCCAAAA 1926
Qy 2085 CTGAATTTCTAGCCGCCAGAGGTTACGAAATTTAGTTAGTCCCATAAACACTTTCTCCA 2144
Db 1927 CTGAATTTCTAGCCGCCAGAGGTTACGAAATTTAGTTAGTCCCATAAACACTTTCTCCA 1986
Qy 2145 GAGCGGAAATCTCTTTTGGAGTATTTCCCAAGAAACGCTTCCCGAAGTTTAAGAGA 2204
Db 1987 GAGCGGAAATCTCTTTTGGAGTATTTCCCAAGAAACGCTTCCCGAAGTTTAAGAGA 2046
Qy 2205 CATGATGTTTAACTCAACAGGAGGCCAGGAGGAGGTAATTTCCCTGGTAAATATT 2264
Db 2047 CATGATGTTTAACTCAACAGGAGGCCAGGAGGAGGTAATTTCCCTGGTAAATATT 2106
Qy 2265 TCACCTCTTCCAAAGAGCCGCTGAGTCTCAATACCTTTGGATTTCGAACCTGGAAACAT 2324
Db 2107 TCACCTCTTCCAAAGAGCCGCTGAGTCTCAATACCTTTGGATTTCGAACCTGGAAACAT 2166
Qy 2325 GGAGACATCTTTGAAAGGATACAAATTTGTCAGTCAAGCTTGTGAGATCATTTCTG 2384
Db 2167 GGAGACATCTTTGAAAGGATACAAATTTGTCAGTCAAGCTTGTGAGATCATTTCTG 2226
Qy 2385 ATGAACCTCAGCATGCTTAAATTAATAAATCAAGCTATTAATACAGATGAAACAAATGAC 2444
Db 2227 ATGAACCTCAGCATGCTTAAATTAATAAATCAAGCTATTAATACAGATGAAACAAATGAC 2286

Qy 2445 AGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGGA 2504
Db 2287 AGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGGA 2346
Qy 2505 GTGTCTGAAAGATTTGCAGAGGTTGACTTTTCTGCGAGTGAGTGTAAGAGTGAATGGTCA 2564
Db 2347 GTGTCTGAAAGATTTGCAGAGGTTGACTTTTCTGCGAGTGAGTGTAAGAGTGAATGGTCA 2406
Qy 2565 GACCAGGTCAGAAATCCACCCCTGACCTTGGAGACACAGCAAGATTTAGAGTGGAACT 2624
Db 2407 GACCAGGTCAGAAATCCACCCCTGACCTTGGAGACACAGCAAGATTTAGAGTGGAACT 2466
Qy 2625 CACACCCAAAAACCATAGGCGGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATTTT 2684
Db 2467 CACACCCAAAAACCATAGGCGGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATTTT 2526
Qy 2685 CCACCTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGAAAGAAAAAGAGAACAGT 2744
Db 2527 CCACCTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGAAAGAAAAAGAGAACAGT 2586
Qy 2745 AGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2804
Db 2587 AGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2646
Qy 2805 CTGCGAGCTTACACAGATAGTTACTTTGGGCTTTTGGCATTTGGCATGGGAGAAAAAGTATTT 2864
Db 2647 CTGCGAGCTTACACAGATAGTTACTTTGGGCTTTTGGCATTTGGCATGGGAGAAAAAGTATTT 2706
Qy 2865 CAAGATCTTCTCGACGAAGAGAGTCAATTGAAGACACAAATTTGGCATACTTCACTGATAGC 2924
Db 2707 CTAGATCTTCTCGACGAAGAGAGTCAATTGAAGACACAAATTTGGCATACTTCACTGATAGC 2766
Qy 2925 AAAATACTGGAGGCAACTAAAGACATATTGAGATTTCCCTCAGATATGTAAATATAA 2984
Db 2767 AAGAATACTGGAGGCAACTAAAGACATATTGAGATTTCCCTCAGATATGTAAATATAA 2826
Qy 2985 ATTCTAAATAGCAAGTTTGGATTTCATCGGGGAAGTCCCTGCTCACATGCTCCACATG 3044
Db 2827 ATTCTAAATAGCAAGTTTGGATTTCATCGGGGAAGTCCCTGCTCACATGCTCCACATG 2886
Qy 3045 ATTGACCGGATTTATGCAAGAACTGCAAGATATGTTTCCCTGAAGAAATTTGCAAGAGC 3104
Db 2887 ATTGACCGGATTTATGCAAGAACTGCAAGATATGTTTCCCTGAAGAAATTTGCAAGAGC 2946
Qy 3105 TCATTTCAAAAGTGGCGCATCTCAGGATATGAGTTTGGCTTCTCTTATTTTATTTAT 3164
Db 2947 TCATTTCAAAAGTGGCGCATCTCAGGATATGAGTTTGGCTTCTCTTATTTTATTTAT 3006
Qy 3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224
Db 3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066
Qy 3225 CAATCTGTGCTTGTCTGACAGAGAAATCCGAACTGGCTACAGAAATTCAGAACTG 3284
Db 3067 CAATCTGTGCTTGTCTGACAGAGAAATCCGAACTGGCTACAGAAATTCAGAACTG 3126
Qy 3285 CCGTTTAAAGTTTGCAGGATTTGACAGGCTTGGAAACACATGCTTAATAAATTTGCTCAAAAAATG 3344
Db 3127 CCGTTTAAAGTTTGCAGGATTTGACAGGCTTGGAAACACATGCTTAATAAATTTGCTCAAAAAATG 3186
Qy 3345 CTTCTGTGTATATCACGAGCTAAATATATTTCCCAACTCAGGAATCTCTACTATGAT 3404
Db 3187 CTTCTGTGTATATCACGAGCTAAATATATTTCCCAACTCAGGAATCTCTACTATGAT 3246
Qy 3405 CCCAACCTTCCCAAGGTCCTGAAAGTCTAGTAACAACTGTAAACCAAGTAACTGACAAA 3464
Db 3247 CCCAACCTTCCCAAGGTCCTGAAAGTCTAGTAACAACTGTAAACCAAGTAACTGACAAA 3306
Qy 3465 ATCCCAAGGATATTAAGGACAAAAAATAATATAGTTTGAATCATGGGAGAGAGAA 3524
Db 3307 ATCCCAAGGATATTAAGGACAAAAAATAATATAGTTTGAATCATGGGAGAGAGAA 3366

QY 3525 ATCGCTTTTAAATGATTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATA 3584
DB 3367 ATCGCTTTTAAATGATTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATA 3426
QY 3585 AGAAAAAACCCTAGGAAGTTTGTTCCTGTAATGACAAACATGACCAACATCAAAAGAT 3644
DB 3427 AGAAAAAACCCTAGGAAGTTTGTTCCTGTAATGACAAACATGACCAACATCAAAAGAT 3486
QY 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3704
DB 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3546
QY 3705 CAATTTGAACCTCCAGAGAGATCGAAACCGTTTCCCTTCATATGATGAGCTGCGAGAA 3764
DB 3547 CAATTTGAACCTCCAGAGAGATCGAAACCGTTTCCCTTCATATGATGAGCTGCGAGAA 3606
QY 3765 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTTACTAGCAACATTTGAT 3824
DB 3607 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTTACTAGCAACATTTGAT 3666
QY 3825 ATGTTTACTATATTCTCATTTTCTGAGCAGTTTAAATGACCTTAAGCGGAAGATTTT 3884
DB 3667 ATGTTTACTATATTCTCATTTTCTGAGCAGTTTAAATGACCTTAAGCGGAAGATTTT 3726
QY 3885 CCCAGAAGGAGATACAAAGAGATAGTCCCAATCGAATCAGAGTATAGAAGATC 3941
DB 3727 CCCAGAAGGAGATACAAAGAGATAGTCCCAATCGAATCAGAGTATAGAAGATC 3783

RESULT 8

US-09-636-060C-20
; Sequence 20, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-060C-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 165 ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCTGCCACAGGTATGGGCTC 224
DB 7 ATGGGGTTCAAGCTCTTGCAGAGACAAACCTATACCTGCTGCCACAGGTATGGGCTC 66
QY 225 TACGTGTGCTTCTGGGGCTGCTGTGTCACCAATCGCTCTCGGCTTCCAGTTCGGAGAGTG 284
DB 67 TACGTGTGCTTCTGGGGCTGCTGTGTCACCAATCGCTCTCGGCTTCCAGTTCGGAGAGTG 126
QY 285 GTTCTGGATGAGCGCGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATT 344
DB 127 GTTCTGGATGAGCGCGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATT 186
QY 345 GCTGGAAGTCCCTTTTCAAGATCGGCTTTGTCTGCCCATCGCGAATGAGCTGTTTACACC 404
DB 187 GCTGGAAGTCCCTTTTCAAGATCGGCTTTGTCTGCCCATCGCGAATGAGCTGTTTACACC 246
QY 405 TGGGTGAATGSCACAGATCTTGAATCTAGAGGAATACAGCAGGTGAGAGACAGATG 464
DB 247 TGGGTGAATGSCACAGATCTTGAATCTAGAGGAATACAGCAGGTGAGAGACAGATG 306

QY 465 GAGGAGGAGCAGAAAGCAATGAGAGAAATCCTTGGGAAAAACACAAAGGAAACCTACTAAG 524
DB 307 GAGGAGGAGCAGAAAGCAATGAGAGAAATCCTTGGGAAAAACACAAAGGAAACCTACTAAG 366
QY 525 AAGAGTGAGAAAGCAGATTAGAGTGTCTTAAACACACTGCAATTAAGGTGCAATGCTTGA 584
DB 367 AAGAGTGAGAAAGCAGATTAGAGTGTCTTAAACACACTGCAATTAAGGTGCAATGCTTGA 426
QY 585 CTGACCCAGCCCTGCGAGCAACATCACCTGGAAGGAGTGCCATCTCTTTATCTCTTCT 644
DB 427 CTGACCCAGCCCTGCGAGCAACATCACCTGGAAGGAGTGCCATCTCTTTATCTCTTCT 486
QY 645 TTTCAATCTCCAGTGCATTTTCAATGTTGAAAAACAAAAACCCCTTCTACCAATGTC 704
DB 487 TTTCAATCTCCAGTGCATTTTCAATGTTGAAAAACAAAAACCCCTTCTACCAATGTC 546
QY 705 TCAGTTGTTGTTTGTGACAGTACTAAGGATGTTGAAGATGCCACATCTCTGAGCTGTAA 764
DB 547 TCAGTTGTTGTTTGTGACAGTACTAAGGATGTTGAAGATGCCACATCTCTGAGCTGTAA 606
QY 765 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGATATAAGAGTCCCTGGA 824
DB 607 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGATATAAGAGTCCCTGGA 666
QY 825 TTAGTCTAATGCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGAGAAACA 884
DB 667 TTAGTCTAATGCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGAGAAACA 726
QY 885 AATCACTAAAAACAAAATGCGAGAAATCTTCTCTAAAGTCAACGTGTGCAAGTTG 944
DB 727 AATCACTAAAAACAAAATGCGAGAAATCTTCTCTAAAGTCAACGTGTGCAAGTTG 786
QY 945 TATTTCAGAGCCAGTGTAGCGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG 1004
DB 787 TATTTCAGAGCCAGTGTAGCGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG 846
QY 1005 AATAAGCAAACTAAGAAACATGACCATGTATGAAAAAGAACTGACATAGTCTGCA 1064
DB 847 AATAAGCAAACTAAGAAACATGACCATGTATGAAAAAGAACTGACATAGTCTGCA 906
QY 1065 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124
DB 907 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966
QY 1125 AGTCGTTTGAAGATAACGAAGAACTGAGTACTCATTCGGATCTATCGAGAGGCAATGCA 1184
DB 967 AGTCGTTTGAAGATAACGAAGAACTGAGTACTCATTCGGATCTATCGAGAGGCAATGCA 1026
QY 1185 CCATGGGTTCCGGAATTTTTCATTTGTCACCAAGGAGATTCATCTCCCTGGCTGAACTT 1244
DB 1027 CCATGGGTTCCGGAATTTTTCATTTGTCACCAAGGAGATTCATCTCCCTGGCTGAACTT 1086
QY 1245 GACAACTCCTCGAGTGAACATAGTAAACACACAGGATGTTTTTGGAAATTTGAGCCACTTG 1304
DB 1087 GACAACTCCTCGAGTGAACATAGTAAACACACAGGATGTTTTTGGAAATTTGAGCCACTTG 1146
QY 1305 CCTACCTTTAGTTTCACTGCTATTGAAATGACATTCATCGGATCGAAGGCTGTCCAG 1364
DB 1147 CCTACCTTTAGTTTCACTGCTATTGAAAGTCACTGTCATCGCATCGAAGGCTGTCCAG 1206
QY 1365 AAGTTTATTACCTAATGATGATGTCATGTTTGGGAAGGATGCTGTGGCAGATGATTTT 1424
DB 1207 AAGTTTATTACCTAATGATGATGTCATGTTTGGGAAGGATGCTGTGGCAGATGATTTT 1266
QY 1425 TACAGTCACTCCAAAGCCAGAAAGTTTATTGACATGGCCTGTGCCAACTGTGCGGAG 1484
DB 1267 TACAGTCACTCCAAAGCCAGAAAGTTTATTGACATGGCCTGTGCCAACTGTGCGGAG 1326
QY 1485 GGCTGCCAGGTTCTCGGATTAAGGATGGCTATTGTGACAAAGCTTGTATTAATTCAGCC 1544
DB 1327 GGCTGCCAGGTTCTCGGATTAAGGATGGCTATTGTGACAAAGCTTGTATTAATTCAGCC 1386
QY 1545 TGCGATTGGGATGGTGGGGATTGCTCTCTGGAACAGTGGAGGGAGTCCGTATATTGAGGA 1604

1387 TCGGATTTGGGATGGTGGGATTTGCTCTGGAAACAGTGGAGGAGTTCGCTATTTTCAGGA 1446
1605 GGTGGAGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC 1664
1447 GGTGGAGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC 1506
1665 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATGATGTTCTGTGAC 1724
1507 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATGATGTTCTGTGAC 1566
1725 CAAGCATCAATGTCCTGCTGTGGGTTTGATGTCGCGACTGTGGGCAAGATCAATTT 1784
1567 CAAGCATCAATGTCCTGCTGTGGGTTTGATGTCGCGACTGTGGGCAAGATCAATTT 1626
1785 CATGAATTTGTAAGTGCATCTTCTCCAAACAGACTCACTATATTTATTCGAAAAGGT 1844
1627 CATGAATTTGTAAGTGCATCTTCTCCAAACAGACTCACTATATTTATTCGAAAAGGT 1686
1845 GAATGCTTGCCTTATTTTCAGCTTTTCAGAACTAGCCAAAGAGGAGTTGAAGTGCCTAT 1904
1687 GAATGCTTGCCTTATTTTCAGCTTTTCAGAACTAGCCAAAGAGGAGTTGAAGTGCCTAT 1746
1905 AGTGCAATCCAAATTTTCAGCTTTTCAGAACTAGCCAAAGAGGAGTTGAAGTGCCTAT 1964
1747 AGTGCAATCCAAATTTTCAGCTTTTCAGAACTAGCCAAAGAGGAGTTGAAGTGCCTAT 1806
1965 ATAATGCAAGTGGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 2024
1807 ATAATGCAAGTGGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1866
2025 AACGATGAAGTTCGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 2084
1867 AACGATGAAGTTCGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1926
2085 CTGATTTCTAGGCGCCAGAGGGTTAGCAAAATTTAGTTAGTCCCATTAACACATCTCTTCCA 2144
1927 CTGATTTCTAGGCGCCAGAGGGTTAGCAAAATTTAGTTAGTCCCATTAACACATCTCTTCCA 1986
2145 GAGCGGGAATTCCTTTTGGAGATATCCCAAGAAACACGCTTCCCGAAGTTTAAAGAGA 2204
1987 GAGCGGGAATTCCTTTTGGAGATATCCCAAGAAACACGCTTCCCGAAGTTTAAAGAGA 2046
2205 CATGATGTTAACTCAACAGAGAGAGCCAGAGAGAGTGAATAATTCCTCGTGAATATT 2264
2047 CATGATGTTAACTCAACAGAGAGAGCCAGAGAGAGTGAATAATTCCTCGTGAATATT 2106
2265 TCACCTCTTCCAAAGAGAGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACAT 2324
2107 TCACCTCTTCCAAAGAGAGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACAT 2166
2325 GGAGCATCATCTTGAAGAGTACAAATTTGTCAGAGTCCGCTTGTGAGATCATTTCTG 2384
2167 GGAGCATCATCTTGAAGAGTACAAATTTGTCAGAGTCCGCTTGTGAGATCATTTCTG 2226
2385 ATGAACTCACAGCATGCTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGAC 2444
2227 ATGAACTCACAGCATGCTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGAC 2286
2445 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAATAAAGCATCTTTGCCAAACAGCTTAGGA 2504
2287 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAATAAAGCATCTTTGCCAAACAGCTTAGGA 2346
2505 GTGCTGAAGATTCAGAGTTCGCTTTCTCCAGTGTGATGTAAGTCAATGGTTCAT 2564
2347 GTGCTGAAGATTCAGAGTTCGCTTTCTCCAGTGTGATGTAAGTCAATGGTTCAT 2406
2565 GACCAAGGTTCAGATTCACCCCTGGACTTGGAGACCCAGCAAGATTTAGAGTGGAAACT 2624
2407 GACCAAGGTTCAGATTCACCCCTGGACTTGGAGACCCAGCAAGATTTAGAGTGGAAACT 2466
2625 CACACCCCAAAACCATAGGCGGAATGTGCAAAAGAGCCCTCTCTGATGTT 2684

2467 CACACCCCAAAACCATAGGCGGAATGTGCAAAAGAAAGCCCTCTCTGATTGTT 2526
2685 CCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGT 2744
2527 CCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGT 2586
2745 AGAATCGAGGAAATGCTGAAATCACATAGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2804
2587 AGAATCGAGGAAATGCTGAAATCACATAGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2646
2805 CTGAGCATTAACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAGATTTTC 2864
2647 CTGAGCATTAACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAGATTTTC 2706
2865 CAAGATCTTCTCGAGAGAGAGGATTCATTTGAAGACACAAATTTGGCATCTTCACTATAGC 2924
2707 CTAGATCTTCTCGAGAGAGAGGATTCATTTGAAGACACAAATTTGGCATCTTCACTATAGC 2766
2925 AAAAACTACTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2984
2767 AAAAACTACTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2826
2985 ATTCTAATAGCAAGTTTGGATTCATCGCGGAAGTCCCTGCTCAGATGCTCCTCACATG 3044
2827 ATTCTAATAGCAAGTTTGGATTCATCGCGGAAGTCCCTGCTCAGATGCTCCTCACATG 2886
3045 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGACG 3104
2887 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGACG 2946
3105 TCATTTCAAAAGTGGGCAATTCAGGATATGAGTTTGGCTTCTCTATTTTATTTAT 3164
2947 TCATTTCAAAAGTGGGCAATTCAGGATATGAGTTTGGCTTCTCTATTTTATTTAT 3006
3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3224
3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3066
3225 CAATCTGGTGTCTTGTCTGACAGAGAAATCGGAACTCTGGCTACAGAAATTCAGAACTG 3284
3067 CAATCTGGTGTCTTGTCTGACAGAGAAATCGGAACTCTGGCTACAGAAATTCAGAACTG 3126
3285 CGGTTAAGTTTCGAGATTTGACAGTCTGGAGACATGCTAAATAATTTGCTCAAAAATG 3344
3127 CGGTTAAGTTTCGAGATTTGACAGTCTGGAGACATGCTAAATAATTTGCTCAAAAATG 3186
3345 CTTCTCTGTGATATCACGAGCTAAATAATTTCCACCAACTCAGGAACTCTTACTATGAT 3404
3187 CTTCTCTGTGATATCACGAGCTAAATAATTTCCACCAACTCAGGAACTCTTACTATGAT 3246
3405 CCCAAGCTGCAAGCTTCACTAAAGTCTAGTAAACAACTGTAAACCACTGACACAA 3464
3247 CCCAAGCTGCAAGCTTCACTAAAGTCTAGTAAACAACTGTAAACCACTGACACAA 3306
3465 ATCCCAAGAGCATAAAGACAAACAAATATAGTTTGAATCATGCGGAGAAAGAA 3524
3307 ATCCCAAGAGCATAAAGACAAACAAATATAGTTTGAATCATGCGGAGAAAGAA 3366
3525 ATGCGCTTTTAAATGATTCGACAAAGCTTCTCATGTTGGTGGCCAGTTGGATGACATA 3584
3367 ATGCGCTTTTAAATGATTCGTAACCAAGCTTCTCATGTTGGTGGCCAGTTGGATGACATA 3426
3585 AGAAAAACCCCTAGCAAGTTTGTTCCTGAAATGCAACAACTGACCAACATCATAAAGAT 3644
3427 AGAAAAACCCCTAGCAAGTTTGTTCCTGAAATGCAACAACTGACCAACATCATAAAGAT 3486
3645 GTCAGACAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3704
3487 GTCAGACAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3546
3705 CAATTTGAACCTGCCAAGAGATTCGAAACCGTTTCTTCTATATGATGATGACCTGAGAA 3764
3547 CAATTTGAACCTGCCAAGAGATTCGAAACCGTTTCTTCTATATGATGATGACCTGAGAA 3606

RESULT 10
US-09-636-596C-20
; Sequence 20, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHOLIPSTER ALPHA-GLUCNcase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-596C-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy	165	ATGCTGTTCAAGCTCTCTGCAGAGACAAACCTATACCTGCTGCTCCACAGGTATGGGCTC	224
Db	7	ATGGGGTTCAAGCTCTTGCAGAGACAAACCTATACCTGCTGCTCCACAGGTATGGGCTC	66
Qy	225	TACGTGTGCTCTTGGGCGTCTGTCACCATCGTCTCGGCTTCCAGTTCGGAGAGTG	284
Db	67	TACGTGTGCTCTTGGGCGTCTGTCACCATCGTCTCGGCTTCCAGTTCGGAGAGTG	126
Qy	285	GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGATTCCTATAGACAAATATT	344
Db	127	GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGATTCCTATAGACAAATATT	186
Qy	345	GCTGGAAGTCTTTTCAAGATCGGTTTGTCTGCCCATGCCGATGACGTTGTTTACACC	404
Db	187	GCTGGAAGTCTTTTCAAGATCGGTTTGTCTGCCCATGCCGATGACGTTGTTTACACC	246
Qy	405	TGGTGATATGGACAGATCTTGAATCTAGTAGAGGAATACACAGGTGAGAGACAGATG	464
Db	247	TGGTGATATGGACAGATCTTGAATCTAGTAGAGGAATACACAGGTGAGAGACAGATG	306
Qy	465	GAGGAGGACAGAAAGCAATGAGAAATCTTGGGAAAAACACACAGCACTACTAAG	524
Db	307	GAGGAGGACAGAAAGCAATGAGAAATCTTGGGAAAAACACACAGCACTACTAAG	366
Qy	525	AAGAGTGAGAGCAGGTAGAGTGTGCTTAAACACACTGCATTAAGGTGCAATGCTTGA	584
Db	367	AAGAGTGAGAGCAGGTAGAGTGTGCTTAAACACACTGCATTAAGGTGCAATGCTTGC	426
Qy	585	CTGGACCCAGCCTGCGCAGCAACATCACCTGAGGAGTGCCATCTCTTTATCCTTCT	644
Db	427	CTGGACCCAGCCTGCGCAGCAACATCACCTGAGGAGTGCCATCTCTTTATCCTTCT	486
Qy	645	TTTCAATCTGCCAGTGACATTTTCAATGTGCAAAAACCAAAAACCCCTTACCAATGTC	704
Db	487	TTTCAATCTGCCAGTGACATTTTCAATGTGCAAAAACCAAAAACCCCTTACCAATGTC	546
Qy	705	TGAGTTGTTGTTTTCAGAGTACTAAGGATGTTGAGATGCCACCTCTGGACTGCTTAA	764
Db	547	TCAGTTGTTGTTTTCAGAGTACTAAGGATGTTGAGATGCCACCTCTGGACTGCTTAA	606
Qy	765	GGAAATAGCAGACAGACAGTATGGAGGGGTACTTGCACACAGATAAAGAACTCCCTGGA	824
Db	607	GGAAATAGCAGACAGACAGTATGGAGGGGTACTTGCACACAGATAAAGAACTCCCTGGA	666
Qy	825	TTAGTCTTAATGCAAGATTGGCTTTCCCTGAGTGGATTTCCACCAACATCAAGGAACA	884
Db	667	TTAGTCTTAATGCAAGATTGGCTTTCCCTGAGTGGATTTCCACCAACATCAAGGAACA	726

Qy	885	AATCAACTAAAAAACAATTCAGAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTTG	944
Db	727	AATCAACTAAAAAACAATTCAGAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTTG	786
Qy	945	TATTCAGAGCCAGTGTAGCGCTTTCTAAAACTCAATAACCCCAAGGATTTTCAAGAA	1004
Db	787	TATTCAGAGCCAGTGTAGCGCTTTCTAAAACTCAATAACCCCAAGGATTTTCAAGAA	846
Qy	1005	AATAAGCAAACTAAGAAGAACATGACATGTATGGAAGAACTGACCATTAAGTCTGCA	1064
Db	847	AATAAGCAAACTAAGAAGAACATGACATGTATGGAAGAACTGACCATTAAGTCTGCA	906
Qy	1065	TATTTATTATGGGATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGC	1124
Db	907	TATTTATTATGGGATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGC	966
Qy	1125	AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATTCGAGAGCGATG	1184
Db	967	AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATTCGAGAGCGATG	1026
Qy	1185	CCATGGGTTCCGGAATATTTTTCATTTGCCAACCGGAGATTCATCTCTGGCTGAACCTT	1244
Db	1027	CCATGGGTTCCGGAATATTTTTCATTTGCCAACCGGAGATTCATCTCTGGCTGAACCTT	1086
Qy	1245	GACAACTCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTTCGAAATTTGAGCCACTT	1304
Db	1087	GACAACTCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTTCGAAATTTGAGCCACTT	1146
Qy	1305	CTTACCTTTAGTTCACCTGCTATTTGAAGTCAATTCATCGCATCGAAGGCTGTCCAG	1364
Db	1147	CTTACCTTTAGTTCACCTGCTATTTGAAGTCAATTCATCGCATCGAAGGCTGTCCAG	1206
Qy	1365	AAGTTTATTACCTTAATGATGATGTCATGTTTGGGAAAGGATGCTGTGCCAGATGATTTT	1424
Db	1207	AAGTTTATTACCTTAATGATGATGTCATGTTTGGGAAAGGATGCTGTGCCAGATGATTTT	1266
Qy	1425	TACAGTCACTCCAAAGGCCAGAAAGTTTATTTGACATGGCTGTGCCAAACTGTGCCGAG	1484
Db	1267	TACAGTCACTCCAAAGGCCAGAAAGTTTATTTGACATGGCTGTGCCAAACTGTGCCGAG	1326
Qy	1485	GGCTGCCAGCTTCCCTGATTAAGGATGCTATTTGTGACAAAGCTGTGTAATTAATTCAGCC	1544
Db	1327	GGCTGCCAGCTTCCCTGATTAAGGATGCTATTTGTGACAAAGCTGTGTAATTAATTCAGCC	1386
Qy	1545	TGCGAATTTGGATGGTGGGATTCCTCTGGAAAACAGTGGAGGGAGTGCCTATATTCAGGA	1604
Db	1387	TGCGAATTTGGATGGTGGGATTCCTCTGGAAAACAGTGGAGGGAGTGCCTATATTCAGGA	1446
Qy	1605	GCTGGAGTACTGGGAGTATTTGGAGTTGGACACCCCTGCGAGTTTGGTGAGGAATAAC	1664
Db	1447	GCTGGAGTACTGGGAGTATTTGGAGTTGGACACCCCTGCGAGTTTGGTGAGGAATAAC	1506
Qy	1665	AGTCTCTTACTGTATTAAGTGATCTTCTCCAAAACAGACTCACTATATTTTCCAAAAGGT	1724
Db	1507	AGTCTCTTACTGTATTAAGTGATCTTCTCCAAAACAGACTCACTATATTTTCCAAAAGGT	1566
Qy	1725	CAAGCATGCAATGCTTGTCTGTGGGTTTGTGCTGGCGACTGTGGGCAAGATCATTTT	1784
Db	1567	CAAGCATGCAATGCTTGTCTGTGGGTTTGTGCTGGCGACTGTGGGCAAGATCATTTT	1626
Qy	1785	CATCAATTTGATTAAGTGATCTTCTCCAAAACAGACTCACTATATTTTCCAAAAGGT	1844
Db	1627	CATCAATTTGATTAAGTGATCTTCTCCAAAACAGACTCACTATATTTTCCAAAAGGT	1686
Qy	1845	GAATGCTGCTGCTTATTTTTCAGCTTTGAGAGTGGCCAAAAGAGAGTTGAAGTGCCTAT	1904
Db	1687	GAATGCTGCTGCTTATTTTTCAGCTTTGAGAGTGGCCAAAAGAGAGTTGAAGTGCCTAT	1746
Qy	1905	AGTCACAACTCCAAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAACCATCCACCTC	1964
Db	1747	AGTCACAACTCCAAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAACCATCCACCTC	1806
Qy	1965	ATAATGCACAGTGGAAATGCAATGCAACCAATACATTTTAACTTCAGTTTCAAAATACA	2024


```
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-635-872A-21

Query Match      62.3%; Score 3485.4; DB 4; Length 3621;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATTTGCTGGAAAGTCC 356
Db 115 AGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAATATTTGCTGGAAAGTCC 174

QY 357 TTTGAGATCGGCTTTGTCGCCATGCGCATGCGATTGACGTTGTTACACCTGGGTGAATGGC 416
Db 175 TTTGAGATCGGCTTTGTCGCCATGCGCATGCGATTGACGTTGTTACACCTGGGTGAATGGC 234

QY 417 ACAGATCTTTGAACTACTCAAGAACTACAGAGGTACAGAGTACAGATGAGAGAGAGCAG 476
Db 235 ACAGATCTTTGAACTACTCAAGAACTACAGAGGTACAGAGTACAGATGAGAGAGAGCAG 294

QY 477 AAAGCAATGAGAGAAATCCTTGGGAAACACACCGGAACCTACTAAGAAAGTGAAG 536
Db 295 AAAGCAATGAGAGAAATCCTTGGGAAACACACCGGAACCTACTAAGAAAGTGAAG 354

QY 537 CAGTTAGAGTGTTCCTAACACACCTGCAATTAAGTGGCCAAATGCTTGTACTGACCCAGCC 596
Db 355 CAGTTAGAGTGTTCCTAACACACCTGCAATTAAGTGGCCAAATGCTTGTACTGACCCAGCC 414

QY 597 CTGCCAGCAACATCACCTGAAGACGTGCGCATCTTTATCTCTTTTCAATCTGCC 656
Db 415 CTGCCAGCAACATCACCTGAAGACGTGCGCATCTTTATCTCTTTTCAATCTGCC 474

QY 657 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCTCAGTTGTTGTT 716
Db 475 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCTCAGTTGTTGTT 534

QY 717 TTTCAGTACTAAGATGTTGAAGATGCCACTCTGCACTGCTTTAAAGGAAATAGCAGA 776
Db 535 TTTCAGTACTAAGATGTTGAAGATGCCACTCTGCACTGCTTTAAAGGAAATAGCAGA 594

QY 777 CAGACAGTATGAGGGGTACTTTGACACAGATAAAGATCCCTGGATTAGTCTAATG 836
Db 595 CAGACAGTATGAGGGGTACTTTGACACAGATAAAGATCCCTGGATTAGTCTAATG 654

QY 837 CAAGATTTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAAA 896
Db 655 CAAGATTTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAAA 714

QY 897 ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCACTGTTATTCAGAGGCC 956
Db 715 ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCACTGTTATTCAGAGGCC 774

QY 957 AGTTAGCGCTTTTAAACTGAATTAACCCCAAGGATTTTCAGAAATTCGAATTAAGCAAACT 1016
Db 775 AGTTAGCGCTTTTAAACTGAATTAACCCCAAGGATTTTCAGAAATTCGAATTAAGCAAACT 834

QY 1017 AAGAAGAAATGACCAATGATGGAAGAACTGACCAATAGTCCCTGCACTATTATTAATGG 1076
Db 835 AAGAAGAAATGACCAATGATGGAAGAACTGACCAATAGTCCCTGCACTATTATTAATGG 894

QY 1077 GATCTGAGCGCATGAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTGA 1136
Db 895 GATCTGAGCGCATGAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTGA 954

QY 1137 GATAACGAAGAACTGAGGTACTCAITGCGATCTATCGAGAGGCGATGCAATGGGTTCCG 1196
Db 955 GATAACGAAGAACTGAGGTACTCAITGCGATCTATCGAGAGGCGATGCAATGGGTTCCG 1014

QY 1197 AATATTTTCAATGTCACCAAGCGGAGATTCATCTCTGCTGAACCTTGACAACTCTCGA 1256
Db 1015 AATATTTTCAATGTCACCAAGCGGAGATTCATCTCTGCTGAACCTTGACAACTCTCGA 1074
```


QY	2337	TTGAAAGGATACAAATTTGTCTCAAGTCAGCCTTGTCTGAGATCATTTCTGATGAACCTCACAG	2399
DB	2155	TTGAAAGGATACAAATTTGTCTCAAGTCAGCCTTGTCTGAGATCATTTCTGATGAACCTCACAG	2214
QY	2397	CATGCTAAATATAAAATTCAAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT	2456
DB	2215	CATGCTAAATATAAAATTCAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT	2274
QY	2457	CCACAGGAAAAACAGGTTTCATAAAGACATCTTGCACAAACAGCTTTAGAGTGCTCTGAAAGA	2516
DB	2275	CCACAGGAAAAACAGGTTTCATAAAGACATCTTGCACAAACAGCTTTAGAGTGCTCTGAAAGA	2334
QY	2517	TTGCAGAGGTTGACTTTTCTCTGCAAGTCAGTGTAAAGTGAATGCTCATGCCAGGCTCAG	2576
DB	2335	TTGCAGAGGTTGACTTTTCTCTGCAAGTCAGTGTAAAGTGAATGCTCATGCCAGGCTCAG	2394
QY	2577	AATCCACCCCTGGACTTTGGAGACACACGCAAGATTTTAGAGTGGAAACTCAACCCCAAAA	2636
DB	2395	AATCCACCCCTGGACTTTGGAGACACACGCAAGATTTTAGAGTGGAAACTCAACCCCAAAA	2454
QY	2637	ACCATAGCGGAATGTGCACAAAGAAAAGCCCCCATCTCTGATTTGTTTCCACTTGGAAAGC	2696
DB	2455	ACCATAGCGGAATGTGCACAAAGAAAAGCCCCCATCTCTGATTTGTTTCCACTTGGAAAGC	2514
QY	2697	CAGATGACAAAAGAAAAAGAAATCACAGGGAAAGAAAAGAGAACAGTGAATCGGAGAA	2756
DB	2515	CAGATGACAAAAGAAAAAGAAATCACAGGGAAAGAAAAGAGAACAGTGAATCGGAGAA	2574
QY	2757	AATGCTGAAAATCACATAGGCGTTACTGAAAGTGTTACTTGGAGAAAGAGCTGCAGCATTTAC	2816
DB	2575	AATGCTGAAAATCACATAGGCGTTACTGAAAGTGTTACTTGGAGAAAGAGCTGCAGCATTTAC	2634
QY	2817	ACAGATAGTTACTTTGGGCTTTTTCGCCATGGAGAAAAAAAGATATTTCCAAGATCTTCTC	2876
DB	2635	ACAGATAGTTACTTTGGGCTTTTTCGCCATGGAGAAAAAAAGATATTTCCTAGATCTTCTC	2694
QY	2877	GACGAAGAAGTCAITTAAGAGACACAATTTGGCATPACTTCACTGATAGCAAAAAATCTGGG	2936
DB	2695	GACGAAGAAGTCAITTAAGAGACACAATTTGGCATPACTTCACTGATAGCAAGAATACTGGG	2754
QY	2937	AGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATAATCTCTAAATAGC	2996
DB	2755	AGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATAATCTCTAAATAGC	2814
QY	2997	AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCTCACATGATTTGACCGGATT	3056
DB	2815	AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCTCACATGATTTGACCGGATT	2874
QY	3057	GTTATGCAAGNACTGCAAGATATGTTCCCTGAGAAATTTGACAGAGCGTCATTTACAAA	3116
DB	2875	GTTATGCAAGNACTGCAAGATATGTTCCCTGAGAAATTTGACAGAGCGTCATTTACAAA	2934
QY	3117	GTGCGCATCTCTGAGGATATGCAGTTTGGCCCTCTCTTATTTTATTTATCTCATGAGTGCA	3176
DB	2935	GTGCGCATCTCTGAGGATATGCAGTTTGGCCCTCTCTTATTTTATTTATCTCATGAGTGCA	2994
QY	3177	GTGCGACCTGAATATATCTCAAGCTTTTGATGAAGTTGATACAGATCAATCTGGTGTC	3236
DB	2995	GTGCGACCTGAATATATCTCAAGCTTTTGATGAAGTTGATACAGATCAATCTGGTGTC	3054
QY	3237	TTGTCTCACAGAGAAATCCGAAACACTGGCTACACAGAAATTCACGAATCGCCGTTAAGTTTG	3296
DB	3055	TTGTCTCACAGAGAAATCCGAAACACTGGCTACACAGAAATTCACGAATCGCCGTTAAGTTTG	3114
QY	3297	CAGATTGACAGGCTTGGAAACACATGCTAATAATAATTTGCTCAAAAATGCTTCTGCTGAT	3356
DB	3115	CAGATTGACAGGCTTGGAAACACATGCTAATAATAATTTGCTCAAAAATGCTTCTGCTGAT	3174
QY	3357	ATCACGCAGCTAAATATATTTCACCAACTCAGGAATCCTTACTATGATCCCACTCGCCA	3416
DB	3175	ATCACGCAGCTAAATATATTTCACCAACTCAGGAATCCTTACTATGATCCCACTCGCCA	3234
QY	3417	CCGGTCACTAAAAAGCTAGTAACAAACTGTAAACACAGTAACCTGACAAAAATCCACAAAGA	3476

RESULT 12

```

US-09-636-077A-21
; Sequence 21, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSSOMAL STORAGE DISEASE
; FILE REFERENCE: 1956152US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-077A-21

```

Query Match 62.3%; Score 3485.4; DB 4; Length 3621;
Best Local Similarity 99.7%; Pred. NO. 0;

QY	297	AGCCGAGATCAATACCAATGTTTGTGATCTCTATAGACACAATATTGCTCGAAAGTCC	356
DB	115	AGCCGAGATCAATACCAATGTTTGTGATCTCTATAGACACAATATTGCTCGAAAGTCC	174
QY	357	TTTCAGAATCGCGTTTGTCTGCGCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGC	416
DB	175	TTTCAGAATCGCGTTTGTCTGCGCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGC	234
QY	417	ACAGATCTTGAACCTACTGAAAGAACTACACGAGGTGAGAGAACAGATGGAGAGAGAGCAG	476
DB	235	ACAGATCTTGAACCTACTGAAAGAACTACACGAGGTGAGAGAACAGATGGAGAGAGAGCAG	294
QY	477	AAAGCAATGAGAGAAATCCCTTTGGGAAAAAACAACAACGGAACCTACTAAGAAGAGTGCAGAA	536
DB	295	AAAGCAATGAGAGAAATCCCTTTGGGAAAAAACAACAACGGAACCTACTAAGAAGAGTGCAGAA	354
QY	537	CAGTTAGAGTGTTCGTAAACACACTGCATTAAGGTGCCAATCGTTGTACTGGACCCGCC	596
DB	355	CAGTTAGAGTGTTCGTAAACACACTGCATTAAGGTGCCAATCGTTGTACTGGACCCGCC	414
QY	597	CTGCCGCCAACATCAACCCCTGGAAGGACGTGCCAATCTCTTTATCCCTCTTTTAACTCTGCC	656

Db 415 CTGCAGCAACATCACCTCGAGGACCTGCACATCTCTTTATCCTCTTTTCAITTCGCC 474
Qy 657 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 716
Db 475 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 534
Qy 717 TTGACAGTACTAAGATGTTGAAGATGCCCACTCTGACATGCTTTAAAGGAAATAGCAGA 776
Db 535 TTGACAGTACTAAGATGTTGAAGATGCCCACTCTGACATGCTTTAAAGGAAATAGCAGA 594
Qy 777 CAGACAGTATGAGGGGTACTTGACAAACAGATAAAGAAAGTCCCTGGATAGTCTAATG 836
Db 595 CAGACAGTATGAGGGGTACTTGACAAACAGATAAAGAAAGTCCCTGGATAGTCTAATG 654
Qy 837 CAAGATTTGGCTTCTCAGTGGATTTCCACCAATTTCAAGGAAACAAATCAACTAAAA 896
Db 655 CAAGATTTGGCTTCTCAGTGGATTTCCACCAATTTCAAGGAAACAAATCAACTAAAA 714
Qy 897 ACAAAATTTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTCTGTTTCAAGGCC 956
Db 715 ACAAAATTTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTCTGTTTCAAGGCC 774
Qy 957 AGTGTAGGCTTCTAAACTGATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 1016
Db 775 AGTGTAGGCTTCTAAACTGATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 834
Qy 1017 AAGAGAAATGACCAATTTGATGGAAGAAAGTCAACCAATGCTGCAATTTTATATGG 1076
Db 835 AAGAGAAATGACCAATTTGATGGAAGAAAGTCAACCAATGCTGCAATTTTATATGG 894
Qy 1077 GATCTGAGCGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGA 1136
Db 895 GATCTGAGCGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGA 954
Qy 1137 GATACGAGAACTCAGGACTCATTGCGATCTATCGAGAGGCGATGCAACCATGGTTCGG 1196
Db 955 GATACGAGAACTCAGGACTCATTGCGATCTATCGAGAGGCGATGCAACCATGGTTCGG 1014
Qy 1197 AATATTTTCAATGTCACCAAGGCGAGATTCATCTCGCTGAACCTTTGACAATCTCGA 1256
Db 1015 AATATTTTCAATGTCACCAAGGCGAGATTCATCTCGCTGAACCTTTGACAATCTCGA 1074
Qy 1257 GTGACAAATGATACACACGAGATGTTTTCGAAATTTGAGCACTGCTACCTTAGT 1316
Db 1075 GTGACAAATGATACACACGAGATGTTTTCGAAATTTGAGCACTGCTACCTTAGT 1134
Qy 1317 TCACCTGCTATTGAAAGTCAACATTCATCGCATCGAAGGCTGTCCGAGAGTTTATTATAC 1376
Db 1135 TCACCTGCTATTGAAAGTCAACATTCATCGCATCGAAGGCTGTCCGAGAGTTTATTATAC 1194
Qy 1377 CTAATGATGATGTCATGTTGGGAAAGATGTCGGCCAGATGATTTTACAGTCACTCC 1436
Db 1195 CTAATGATGATGTCATGTTGGGAAAGATGTCGGCCAGATGATTTTACAGTCACTCC 1254
Qy 1437 AAGGCCAAGGTTTATTGACATGGCTGTGCGCAACTGTGCGAGGGCTGCGCCAGGT 1496
Db 1255 AAGGCCAAGGTTTATTGACATGGCTGTGCGCAACTGTGCGAGGGCTGCGCCAGGT 1314
Qy 1497 TCCTGGATTAAGATGCTTATTGACAAAGGCTGTGAATTAATTCAGCCTGCGATGGAT 1556
Db 1315 TCCTGGATTAAGATGCTTATTGACAAAGGCTGTGAATTAATTCAGCCTGCGATGGAT 1374
Qy 1557 GGTGGGATGCTCTGGAACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGGTACT 1616
Db 1375 GGTGGGATGCTCTGGAACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGGTACT 1434
Qy 1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATTAACAGTGTCTCTTAC 1676
Db 1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATTAACAGTGTCTCTTAC 1494
Qy 1677 TGTAAATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTCTTGTGACCAAGCATGCAAT 1736

Db 1495 TGTAAATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTCTTGTGACCAAGCATGCAAT 1554
Qy 1737 GTCTTCTCTCTGTTGTTTTCATGCTGCGACTGTGGCAAGATCATTTTTCATGAATGTAT 1796
Db 1555 GTCTTCTCTCTGTTGTTTTCATGCTGCGACTGTGGCAAGATCATTTTTCATGAATGTAT 1614
Qy 1797 AAAAGTGTATCTTCTCCAAAACAGACTCTATATATTATTTCCAAAAGGTGAATGCCCT 1856
Db 1615 AAAAGTGTATCTTCTCCAAAACAGACTCTATATATTATTTCCAAAAGGTGAATGCCCT 1674
Qy 1857 TATTTTCAGCTTTGCAAGATGACCAAGAGAGGATGAGGTGCTATAGTCACATCCA 1916
Db 1675 TATTTTCAGCTTTGCAAGATGACCAAGAGAGGATGAGGTGCTATAGTCACATCCA 1734
Qy 1917 ATAATTCGACATGCTTCTATTTGCCAAACAGTGGAAAACCATCCACTCTATAATGACAGT 1976
Db 1735 ATAATTCGACATGCTTCTATTTGCCAAACAGTGGAAAACCATCCACTCTATAATGACAGT 1794
Qy 1977 GGAATGAATGCCACCAACATATATTTTATCTCACGTTTCAAAAATACAAACGATGAAG 2036
Db 1795 GGAATGAATGCCACCAACATATATTTTATCTCACGTTTCAAAAATACAAACGATGAAG 1854
Qy 2037 TTCAAAATGCAGATAACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAAATTTCTACG 2096
Db 1855 TTCAAAATGCAGATAACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAAATTTCTACG 1914
Qy 2097 GCCCAGAGGGTTTACGAAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 2156
Db 1915 GCCCAGAGGGTTTACGAAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 1974
Qy 2157 CTTTTCAGGATATTCCTCAAGAAACCGCTTCCCGAAGTTTAAAGAGACATGATGTTAAC 2216
Db 1975 CTTTTCAGGATATTCCTCAAGAAACCGCTTCCCGAAGTTTAAAGAGACATGATGTTAAC 2034
Qy 2217 TCACCAAGGAGAGGCCAGGAAAGAGTGAATAATTTCCCTCTGTTAAATTTTCACTCTCTCCA 2276
Db 2035 TCACCAAGGAGAGGCCAGGAAAGAGTGAATAATTTCCCTCTGTTAAATTTTCACTCTCTCCA 2094
Qy 2277 AAGAGACCCAGTTGATGCTCAATAGCTTTGGATTTGCAACTGGAACTGGAGACATCACT 2336
Db 2095 AAGAGACCCAGTTGATGCTCAATAGCTTTGGATTTGCAACTGGAACTGGAGACATCACT 2154
Qy 2337 TTCAAAAGGATACAAATTTGTCAGAGTCAGCTTGTCTGAGATCATTTCTGATGAACCTCACAG 2396
Db 2155 TTGAAGGATACAAATTTGTCAGAGTCAGCTTGTCTGAGATCATTTCTGATGAACCTCACAG 2214
Qy 2397 CATGCTAAATAAAAAATCAAGCTATAATAACAGATGAACCAAAATGACAGTTTGGTGGCT 2456
Db 2215 CATGCTAAATAAAAAATCAAGCTATAATAACAGATGAACCAAAATGACAGTTTGGTGGCT 2274
Qy 2457 CCACAGGAAACAGGTTTCATTAAGCACTCTTCCCAACAGCTTTAGGAGTGTCTGAAAGA 2516
Db 2275 CCACAGGAAACAGGTTTCATTAAGCACTCTTCCCAACAGCTTTAGGAGTGTCTGAAAGA 2334
Qy 2517 TTGCAGAGTTGACTTTTCTGCTGAGTGAATGTAAGTGAATGCTCATGCCAGGCTCAG 2576
Db 2335 TTGCAGAGTTGACTTTTCTGCTGAGTGAATGTAAGTGAATGCTCATGCCAGGCTCAG 2394
Qy 2577 AATCCACCCCTGACCTGGAGACCCACAGCAGATTTAGATGGAACCTCACACCCAAAA 2636
Db 2395 AATCCACCCCTGACCTGGAGACCCACAGCAGATTTAGATGGAACCTCACACCCAAAA 2454
Qy 2637 ACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC 2696
Db 2455 ACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC 2514
Qy 2697 CAGATGACAAAAGAAAGAAATCAGAGGAAAGAAAGAAAGAGAACAGTAGAATGGAGGAA 2756
Db 2515 CAGATGACAAAAGAAAGAAATCAGAGGAAAGAAAGAAAGAGAACAGTAGAATGGAGGAA 2574
Qy 2757 AATGCTGAAATCAGATAGGCTTACTGAAAGTGTACTTTGAAAGAAAGCTGACAGCATTAC 2816
Db 2575 AATGCTGAAATCAGATAGGCTTACTGAAAGTGTACTTTGAAAGAAAGCTGACAGCATTAC 2634

2817 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGATATTTCCAGATCTTCTC 2876
2635 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGATATTTCTAGATCTTCTC 2694
2877 GACGAGAGAGTCTATTCAGACACAAATGGCATACTTCTCTGATAGCAAAAAATCTCTGG 2936
2695 GACGAGAGAGTCTATTCAGACACAAATGGCATACTTCTCTGATAGCAAAAAATCTCTGG 2754
2937 AGGCAACTAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAAATTTCTAAATAGC 2996
2755 AGGCAACTAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAAATTTCTAAATAGC 2814
2997 AGTTTGGATTACATTCGGGAAAGTCCCTGCTCAGATGCTCAGATGATTCAGCGGAT 3056
2815 AGTTTGGATTACATTCGGGAAAGTCCCTGCTCAGATGCTCAGATGATTCAGCGGAT 2874
3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGCAAGAAATTTGCAAGAGCTCATTTTCAAAA 3116
2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGCAAGAAATTTGCAAGAGCTCATTTTCAAAA 2934
3117 GTGCGCCATTTGAGGATATGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCA 3176
2935 GTGCGCCATTTGAGGATATGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCA 2994
3177 GTGCGCCATTTGAGGATATGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCA 3236
2995 GTGCGCCATTTGAGGATATGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCA 3054
3237 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTTGCAAGAGCTCATTTTCAAAA 3296
3055 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTTGCAAGAGCTCATTTTCAAAA 3114
3297 CAGGATTTGACAGGTTCTGGAACACATGTAATAATTTGCTCAGAAATTTGCAAGAGCTCATTTTCAAAA 3356
3115 CAGGATTTGACAGGTTCTGGAACACATGTAATAATTTGCTCAGAAATTTGCAAGAGCTCATTTTCAAAA 3174
3357 ATCAGCAGCTAAATTAATTTCCACCAACTGAGGAAATTTGCAAGAGCTCATTTTCAAAA 3416
3175 ATCAGCAGCTAAATTAATTTCCACCAACTGAGGAAATTTGCAAGAGCTCATTTTCAAAA 3234
3417 CCGTCTCAATTAAGTCTAGTAACAACTGTAACCAAGTAACTGCAAGAGCTCATTTTCAAAA 3476
3235 CCGTCTCAATTAAGTCTAGTAACAACTGTAACCAAGTAACTGCAAGAGCTCATTTTCAAAA 3294
3477 TATAAGGACAAAAACAATATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAA 3536
3295 TATAAGGACAAAAACAATATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAA 3354
3537 ATGATTTCTGACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATAGAAAAACCCCT 3596
3355 ATGATTTCTGACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATAGAAAAACCCCT 3414
3597 AGGAAGTTTGTTCCTGTAATGACAACTGACCAATTCATAGAGATGCTCAGACAGTG 3656
3415 AGGAAGTTTGTTCCTGTAATGACAACTGACCAATTCATAGAGATGCTCAGACAGTG 3474
3657 AAGCTGTTTCTCAGGAGCTTCTATGATTCATGTTCCCATGCTTCCCAATTTGAACTG 3716
3475 AAGCTGTTTCTCAGGAGCTTCTATGATTCATGTTCCCATGCTTCCCAATTTGAACTG 3534
3717 CCRAGAGATATCAAAACCTTCTCTCATATGATGCTCAGGATGCTCAGGAGCTTAT 3776
3535 CCRAGAGATATCAAAACCTTCTCTCATATGATGCTCAGGATGCTCAGGAGCTTAT 3594
3777 CGAGACAAATGGAAGTTTGGAC 3799
3595 CGAGACAAATGGAAGTAGTAGTC 3617

Patent No. 6642038
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
FILE REFERENCE: 210119US00CONT
CURRENT APPLICATION NUMBER: US/09/636,060C
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 3621
TYPE: DNA
ORGANISM: Homo sapiens
US-09-636-060C-21

Query Match 62.3%; Score 3485.4; DB 4; Length 3621;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

297 ACCGAGATCAATACATATGTTTGTGATTCCTTATAGAGACAATTTGCTGGAAAGTCC 356
115 AGCCGAGATCAATAGCATGTTTGTGATTCCTTATAGAGACAATTTGCTGGAAAGTCC 174
357 TTTTCAAGATCGGCTTTGCTGCCCCATGCCGATGACGTTGTTTACACCTGGGTGAATGC 416
175 TTTTCAAGATCGGCTTTGCTGCCCCATGCCGATGACGTTGTTTACACCTGGGTGAATGC 234
417 ACAGATCTTTGAACTACTGAAGGAACCTACAGAGGTACAGAGCAAGATCGAGAGAGCAG 476
235 ACAGATCTTTGAACTACTGAAGGAACCTACAGAGGTACAGAGCAAGATCGAGAGAGCAG 294
477 AAAGCAATGAGAGAAATCTTTGGGAAACACACAGGAACTTCTTAAAGAGGTGAGAG 536
295 AAAGCAATGAGAGAAATCTTTGGGAAACACACAGGAACTTCTTAAAGAGGTGAGAG 354
537 CAGTTAGAGTGTTCCTTAAACACACTGCAATTAAGGTGCCAATGCTGTACTGGACCCAGCC 596
355 CAGTTAGAGTGTTCCTTAAACACACTGCAATTAAGGTGCCAATGCTGTACTGGACCCAGCC 414
597 CTGCCAGCAACATCACCCTGAGAGAGCTGCACTCTTTATCTCTTTTCTTCTTCTGCCC 656
415 CTGCCAGCAACATCACCCTGAGAGAGCTGCACTCTTTATCTCTTTTCTTCTTCTGCCC 474
657 AGTGACATTTTCAATGTTTGCACAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 716
475 AGTGACATTTTCAATGTTTGCACAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 534
717 TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGA 776
535 TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGA 594
777 CAGACAGTATGAGGGGTACTTGCACACAGATTAAGAGTCCCTGGATAGTCTAATG 836
595 CAGACAGTATGAGGGGTACTTGCACACAGATTAAGAGTCCCTGGATAGTCTAATG 654
837 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACCAAACTCAACTAAA 896
655 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACCAAACTCAACTAAA 714
897 ACAGATTTGCCAGAAATCTTCTCTTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 956
715 ACAGATTTGCCAGAAATCTTCTCTTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 774
957 AGTGTAGCGCTTCTAAAACCTGAATAAACCCTGAGGATTTTCAAGAAATGAATAAGCAAACT 1016
775 AGTGTAGCGCTTCTAAAACCTGAATAAACCCTGAGGATTTTCAAGAAATGAATAAGCAAACT 834
1017 AAGAGAAACATGACCATGATGAGAAAGAACTGACCAATAGTCTGATATTTATG 1076
835 AAGAGAAACATGACCATGATGAGAAAGAACTGACCAATAGTCTGATATTTATG 894

QY	1077	GATCTGAGGCCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTTGAA	1136
DB	895	GATCTGAGCCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTTGAA	954
QY	1137	GATAACGAAGACTGAGTACTCATTCGGATCTATCGAGAGCAGTCACATTCGGTTCGG	1196
DB	955	GATAACGAAGACTGAGTACTCATTCGGATCTATCGAGAGCAGTCACATTCGGTTCGG	1014
QY	1197	AATATTTTCATTTGTCACCAACGGGCAGATTCCTATCCTGGCTGAACCTTGACAATCCTCGA	1256
DB	1015	AATATTTTCATTTGTCACCAACGGGCAGATTCCTATCCTGGCTGAACCTTGACAATCCTCGA	1074
QY	1257	GTGACAAATAGTAACACACACGAGATGTTTTTCGNAATTTGAGCCACTTGCCTACCTTTAGT	1316
DB	1075	GTGACAAATAGTAACACACACGAGATGTTTTTCGNAATTTGAGCCACTTGCCTACCTTTAGT	1134
QY	1317	TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTATAC	1376
DB	1135	TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTATAC	1194
QY	1377	CTAAATGATGATGTCATGTTTTGGGAAGGATGTCGGCCAGATGATTTTTTACAGTCACTCC	1436
DB	1195	CTAAATGATGATGTCATGTTTTGGGAAGGATGTCGGCCAGATGATTTTTTACAGTCACTCC	1254
QY	1437	AAAGGCCAGAAAGTTTATTTGACATGGCTGTGCCCAAACTGTGCCAGAGGCTGCCAGGT	1496
DB	1255	AAAGGCCAGAAAGTTTATTTGACATGGCCCTGTGCCAAACTGTGCCAGAGGCTGCCAGGT	1314
QY	1497	TCCTGGATTAAAGATGGCTATTGTGCAAGGGCTTGTAATAATTCAGCCCTCGATTCGGAT	1556
DB	1315	TCCTGGATTAAAGATGGCTATTGTGCAAGGGCTTGTAATAATTCAGCCCTCGATTCGGAT	1374
QY	1557	GGTGGGATTGCTCTGGARACAGTGGAGGAGTCGCTATATATTCAGAGAGTGGAGTACT	1616
DB	1375	GGTGGGATTGCTCTGGAAACAGTGGAGGAGTCGCTATATTCAGAGAGTGGAGTACT	1434
QY	1617	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC	1676
DB	1435	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC	1494
QY	1677	TGTAATCAGGGATGCGGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATCAAT	1736
DB	1495	TGTAATCAGGGATGCGGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATCAAT	1554
QY	1737	GTCTTGCTCTGGGTTTGATGTGTGGCGACTGTGGCGAAGATCATTTTCATGAATTTGAT	1796
DB	1555	GTCTTGCTCTGGGTTTGATGTGTGGCGACTGTGGCGAAGATCATTTTCATGAATTTGAT	1614
QY	1797	AAAGTGATCCTTCTCCAAACAGACATCACTATATTTCCAAAGGTGAATCGCTGCCT	1856
DB	1615	AAAGTGATCCTTCTCCAAACAGACATCACTATATTTCCAAAGGTGAATCGCTGCCT	1674
QY	1857	TATTTTCAGCTTTGCAGAAGTAGCCAAAGAGGAGTTGAAGGTGCTATAGTGACAATCCA	1916
DB	1675	TATTTTCAGCTTTGCAGAAGTAGCCAAAGAGGAGTTGAAGGTGCTATAGTGACAATCCA	1734
QY	1917	ATAATTCGACATGTTCTATTGCGAACAGTGGAAAAACCATCCACTCATATATGCACAGT	1976
DB	1735	ATAATTCGACATGTTCTATTGCGAACAGTGGAAAAACCATCCACTCATATATGCACAGT	1794
QY	1977	GGATGATGCCACACACATATATTTTAACTCTCAGTTTCCAAATACAAACGATGAAGAG	2036
DB	1795	GGATGATGCCACACACATATATTTTAACTCTCAGTTTCCAAATACAAACGATGAAGAG	1854
QY	2037	TTCAAAATGCAGATAACAGTGGAGTGGACAAAGGAGGGAACAAACCTGAATTTCTACG	2096
DB	1855	TTCAAAATGCAGATAACAGTGGAGTGGACAAAGGAGGGAACAAACCTGAATTTCTACG	1914
QY	2097	GCCAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAACCTTCTTCAGAGGGCGGAATC	2156
DB	1915	GCCAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAACCTTCTTCAGAGGGCGGAATC	1974
QY	2157	CTTTTTTCAGGATATTCCCAAGAAAAACCGTTCTCCGAGTTTAAAGACATGATTTTAAAC	2216

1975	Db	TTTGGAGTATTTCCCAAGAAACCGCTTCCCGAGTTTAAGAGACATGATGTTAC	20334
2217	Qy	TCAACAAGAGAGAGCCCGAGGAAGAGGTGAAATTTCCCTGGTGAATAATTTTCATCTCCTTCCA	2276
2035	Db	TCAACAAGAGAGAGCCCGAGGAAGAGGTGAAATTTCCCTGGTGAATAATTTTCATCTCCTTCCA	2094
2277	Qy	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGAATTTGCAACTGGAAATGAGACATCACT	2336
2095	Db	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGAATTTGCAACTGGAAATGAGACATCACT	2154
2337	Qy	TTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGTCTGAGATCAATTTCTGTAGTAACACAG	2396
2155	Db	TTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGTCTGAGATCAATTTCTGTAGTAACACAG	2214
2397	Qy	CATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAATGACACAGTTTGGTGGCT	2456
2215	Db	CATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAATGACACAGTTTGGTGGCT	2274
2457	Qy	CCACAGGAAACACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGATGTCGAAAGA	2516
2275	Db	CCACAGGAAACACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGATGTCGAAAGA	2334
2517	Qy	TTCCAGAGGTGACTTTTCTGCGAGTGAGTGTAATAAGTGAATGGTCAATGACCAAGGTCAG	2576
2335	Db	TTCCAGAGGTGACTTTTCTGCGAGTGAGTGTAATAAGTGAATGGTCAATGACCAAGGTCAG	2394
2577	Qy	AATCCACCCCTGACCTTGAGAGCCACAGCAAGATTTAGAGTGAACACTCACACCCAAAAA	2636
2395	Db	AATCCACCCCTGACCTTGAGAGCCACAGCAAGATTTAGAGTGAACACTCACACCCAAAAA	2454
2637	Qy	ACCATAGGGCGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATGTTTCCACTGGAAGC	2696
2455	Db	ACCATAGGGCGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATGTTTCCACTGGAAGC	2514
2697	Qy	CAGATCAGAAACAGAAAGAAATACAGGGAAGAAAGAGAAACAGTAGTAATGGAGGAA	2756
2515	Db	CAGATCAGAAACAGAAAGAAATACAGGGAAGAAAGAGAAACAGTAGTAATGGAGGAA	2574
2757	Qy	AATGCTGAAATCACATAGGCGTTACTGAAAGTCTTCTGAAAGAAAGCTGACAGATTAC	2816
2575	Db	AATGCTGAAATCACATAGGCGTTACTGAAAGTCTTCTGAAAGAAAGCTGACAGATTAC	2834
2817	Qy	ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCCAAGATCTTCTC	2876
2635	Db	ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCTAGATCTTCTC	2694
2877	Qy	GACGAAGAGAGTCATTGAAGACACAAATTTGGCATCTTCACTGATGCAAAAAATCTGGG	2936
2695	Db	GACGAAGAGAGTCATTGAAGACACAAATTTGGCATCTTCACTGATGCAAAAAATCTGGG	2754
2937	Qy	AGGCACTAAAGATACATTTGCAGATTCCTCAGATGATGTAATTAATTTCTAAATAGC	2996
2755	Db	AGGCACTAAAGATACATTTGCAGATTCCTCAGATGATGTAATTAATTTCTAAATAGC	2814
2997	Qy	AAGTTTGGATTCACTCGCGGAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATT	3056
2815	Db	AAGTTTGGATTCACTCGCGGAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATT	2874
3057	Qy	GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAAGAGCTCATTTCAAAA	3116
2875	Db	GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAAGAGCTCATTTCAAAA	2934
3117	Qy	GTGCGCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	3176
2935	Db	GTGCGCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	2994
3177	Qy	GTGCGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTC	3236
2995	Db	GTGCGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTC	3054
3237	Qy	TTGTCGCACAGAGAAATCCGAAACACTGGCTACCAAGATTCACGAATTCGGCTTAAGTTTG	3296

2395	DB	AATCCACCCTGGACTTTGGAGACCCACGACGAATTTAGAGTGGAAACTCACACCCAAAAA	2455
2637	QY	ACCATAGGCGGAAATGTGACAAAAGAAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC	2696
2455	DB	ACCATAGGCGGAAATGTGACAAAAGAAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC	2514
2697	QY	CAGATGACAAAGAAAGAAATTCACAGGGAAAGAAABABAGAGAACAGTAGTAATGGAGGAA	2756
2515	DB	CAGATGACAAAGAAAGAAATTCACAGGGAAAGAAABABAGAGAACAGTAGTAATGGAGGAA	2574
2757	QY	AATGCTGAAAAATCACATAGGCGGTTACTTGAAGTGTTTACTTTGGAGAAAGCTGCAGCATTTAC	2816
2575	DB	AATGCTGAAAAATCACATAGGCGGTTACTTGAAGTGTTTACTTTGGAGAAAGCTGCAGCATTTAC	2634
2817	QY	ACAGATAGTTACTTTGGCGTTTTTGGCATGGGAGAAAABAAAAGTTATTTCCAAGATCTTCTC	2876
2635	DB	ACAGATAGTTACTTTGGCGTTTTTGGCATGGGAGAAAABAAAAGTTATTTCCAAGATCTTCTC	2594
2877	QY	GACCAAGAAAGAGTCATTGGAAGACACAATTTGGCATACTTCTCACTGATAGCAAAAATACTGGG	2936
2695	DB	GACCAAGAAAGAGTCATTGGAAGACACAATTTGGCATACTTCTCACTGATAGCAAAAATACTGGG	2754
2937	QY	AGGCAACTAAAGATACATTTGCAGATTCCTTCAGATATGTAAATAAAATCTTAAATAGC	2996
2755	DB	AGGCAACTAAAGATACATTTGCAGATTCCTTCAGATATGTAAATAAAATCTTAAATAGC	2814
2997	QY	AAGTTTGGATTACATCTCGGAAAGTCCCTGCTCCACATGCTCCACATGATGTACCCGGATT	3056
2815	DB	AAGTTTGGATTACATCTCGGAAAGTCCCTGCTCCACATGCTCCACATGATGTACCCGGATT	2874
3057	QY	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAGCGTCAATTCACAAA	3116
2875	DB	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAGCGTCAATTCACAAA	2934
3117	QY	GTGCGCATCTCAGGATATGCAAGTTTGCCCTCTCTTTATTTTATATCTCATGAGTGCA	3176
2935	DB	GTGCGCATCTCAGGATATGCAAGTTTGCCCTCTCTTTATTTTATATCTCATGAGTGCA	2994
3177	QY	GTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGCTGTC	3236
2995	DB	GTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGCTGTC	3054
3237	QY	TTGCTGACAGAGAAATCCGAAACACTGGCTACAGAGATTCACGAACCTGCGTTAAGTTTG	3296
3055	DB	TTGCTGACAGAGAAATCCGAAACACTGGCTACAGAGATTCACGAACCTGCGTTAAGTTTG	3114
3297	QY	CAGGATTTGACAGGCTCTGGAAACATGCTTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT	3356
3115	DB	CAGGATTTGACAGGCTCTGGAAACATGCTTAATAAATTTGCTCAAAAATGCTTCCTGCTGAT	3174
3357	QY	ATCAGCAGCTAAATATATTTCCACCAACTCAGGAATCCCTACTATGATCCCAACCTGCCA	3416
3175	DB	ATCAGCAGCTAAATATATTTCCACCAACTCAGGAATCCCTACTATGATCCCAACCTGCCA	3234
3417	QY	CCGCTCACTAAAAGCTCTAGTAAACAAACCTGTAAACCCAGTAACTGCACAAAATCCACAAAGCA	3476
3235	DB	CCGCTCACTAAAAGCTCTAGTAAACAAACCTGTAAACCCAGTAACTGCACAAAATCCACAAAGCA	3294
3477	QY	TATTAAGGACAAAAACAAATATAGTTTGAATCATGTGGAGAGAGAAATCGCTTTTAAA	3536
3295	DB	TATTAAGGACAAAAACAAATATAGTTTGAATCATGTGGAGAGAGAGAAATCGCTTTTAAA	3354
3537	QY	ATGATTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGATCAATAAGAAAAAACCCCT	3596
3355	DB	ATGATTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGATCAATAAGAAAAAACCCCT	3414
3597	QY	AGGAAGTTTGTTCCTCGAATGACAAATTTGACACACATCATAAAGATGCTTCAGACAGTG	3656
3415	DB	AGGAAGTTTGTTCCTCGAATGACAAATTTGACACACATCATAAAGATGCTTCAGACAGTG	3474
3657	QY	AAAGCTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTTGAACCTG	3716
3475	DB	AAAGCTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTTGAACCTG	3534

QY	1977	GGATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAATACAAACGATGAAG	2036
DB	1795	GGATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAATACAAACGATGAAG	1854
QY	2037	TTCAAAATGCAGATTAACAGTGGAGTGACACAAAGGAGGACCAAAATCTTACG	2096
DB	1855	TTCAAAATGCAGATTAACAGTGGAGTGACACAAAGGAGGACCAAAATCTTACG	1914
QY	2097	GCCAGAGGGTTACGAAATTTAGTTAGTCCATPAACATCTTCCAGAGCGGAAATC	2156
DB	1915	GCCAGAGGGTTACGAAATTTAGTTAGTCCATPAACATCTTCCAGAGCGGAAATC	1974
QY	2157	CTTTTGGAGTATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGAGACATGATTTAA	2216
DB	1975	CTTTTGGAGTATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGAGACATGATTTAA	2034
QY	2217	TCACAAGAGAGCCAGGAAGAGTGAAATTTCCCTGGTAAATATTTCTCTTCCA	2276
DB	2035	TCACAAGAGAGCCAGGAAGAGTGAAATTTCCCTGGTAAATATTTCTCTTCCA	2094
QY	2277	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACTGGAGACATCACT	2336
DB	2095	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACTGGAGACATCACT	2154
QY	2337	TTGAAAGGATACAAATTTGTCCAGTCAGCTTTGCTGAGATCAATTTCTGATCAACTCACAG	2396
DB	2155	TTGAAAGGATACAAATTTGTCCAGTCAGCTTTGCTGAGATCAATTTCTGATCAACTCACAG	2214
QY	2397	CATGCTAAATAAAAATCAAGCTATATPAACAGATGAACAAATGACAGTTTGTGGCT	2456
DB	2215	CATGCTAAATAAAAATCAAGCTATATPAACAGATGAACAAATGACAGTTTGTGGCT	2274
QY	2457	CCACAGGAAACACAGTTTCATAAAGACATCTTCCAAACAGCTTAGGAGTCTCAAGA	2516
DB	2275	CCACAGGAAACACAGTTTCATAAAGACATCTTCCAAACAGCTTAGGAGTCTCAAGA	2334
QY	2517	TTGCAGAGGTTGATTTTCTCAGTGTGCTGATGAAAGTGAATGGTCAAGCAGGGTCAG	2576
DB	2335	TTGCAGAGGTTGATTTTCTCAGTGTGCTGATGAAAGTGAATGGTCAAGCAGGGTCAG	2394
QY	2577	AATCCACCCCTGGACTTGAGAGACACAGCAAGATTTAGAGTGAACCTCACACCCAAA	2636
DB	2395	AATCCACCCCTGGACTTGAGAGACACAGCAAGATTTAGAGTGAACCTCACACCCAAA	2454
QY	2637	ACCATAGGGGAAATGTGACAAAGAAAGCCCATCTCTGATTTGCCACTGGAAAGC	2696
DB	2455	ACCATAGGGGAAATGTGACAAAGAAAGCCCATCTCTGATTTGCCACTGGAAAGC	2514
QY	2697	CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAAACAGTAGAATGAGGAA	2756
DB	2515	CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAAACAGTAGAATGAGGAA	2574
QY	2757	AATGCTGAAATCACATAGCGTTACTGAAAGTGTACTTTGGAAAGAGCTGAGCAATTAC	2816
DB	2575	AATGCTGAAATCACATAGCGTTACTGAAAGTGTACTTTGGAAAGAGCTGAGCAATTAC	2634
QY	2817	ACAGATAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAATTTTCCAGATCTTCTC	2876
DB	2635	ACAGATAGTTACTTGGGCTTTTGGCAATGGGAGAAAGAAATTTTCCAGATCTTCTC	2694
QY	2877	GACGAGAGAGTCAATGAGACACAAATTTGGCATCTTCACTGATAGCAAAATCTGGG	2936
DB	2695	GACGAGAGAGTCAATGAGACACAAATTTGGCATCTTCACTGATAGCAAAATCTGGG	2754
QY	2937	AGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAAATTTCTAAATAGC	2996
DB	2755	AGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAAATTTCTAAATAGC	2814
QY	2997	AAGTTTGGATTCATCGGGAAAGTCCCTGCTCACATGCCCTCACATGATGACCGGATT	3056
DB	2815	AAGTTTGGATTCATCGGGAAAGTCCCTGCTCACATGCCCTCACATGATGACCGGATT	2874
QY	3057	GTTATGCAAGAACTGCAAGATATGTTTCCCTGAAAGATTTGCAAGACGTCATTTCAAAA	3116

DB	2875	GTTATGCAAGAACTGCAAGATATGTTTCCCTGAAAGATTTGCAAGACGTCATTTCAAAA	2934
QY	3117	GTGCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	3176
DB	2935	GTGCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	2994
QY	3177	GTGCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	3236
DB	2995	GTGCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	3054
QY	3237	TTGCTGTACAGAGAAATCCGAAACACCTGGCTACAGAAATTCAGAAATCGCGTTAAATTTG	3296
DB	3055	TTGCTGTACAGAGAAATCCGAAACACCTGGCTACAGAAATTCAGAAATCGCGTTAAATTTG	3114
QY	3297	CAGGATTTGACAGGCTGTGGAACACATCTTAATAAATTTGCTCAAAATGCTTCTCTGCTGAT	3356
DB	3115	CAGGATTTGACAGGCTGTGGAACACATCTTAATAAATTTGCTCAAAATGCTTCTCTGCTGAT	3174
QY	3357	ATCAGCAGCTAAATAATATTTCCACCAACTCAGGAATCTCTTACTATGATCCCAACCTGCCA	3416
DB	3175	ATCAGCAGCTAAATAATATTTCCACCAACTCAGGAATCTCTTACTATGATCCCAACCTGCCA	3234
QY	3417	CCGCTCACTAAAAGTCTAGTAAACAACTGTAAACCAAGTAACTGACAAAATCCCAAGCA	3476
DB	3235	CCGCTCACTAAAAGTCTAGTAAACAACTGTAAACCAAGTAACTGACAAAATCCCAAGCA	3294
QY	3477	TATAAGGACAAAACAAATATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAA	3536
DB	3295	TATAAGGACAAAACAAATATAGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAA	3354
QY	3537	ATGATTCGTACCAACCGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCT	3596
DB	3355	ATGATTCGTACCAACCGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCT	3414
QY	3597	AGGAGTTTGTTCCTGCTGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTG	3656
DB	3415	AGGAGTTTGTTCCTGCTGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTG	3474
QY	3657	AAGGCTCTTCTCAGGGACTTCTATGAATCCATGTTCCTCATACCTTCCCAATTTGAACCTG	3716
DB	3475	AAGGCTCTTCTCAGGGACTTCTATGAATCCATGTTCCTCATACCTTCCCAATTTGAACCTG	3534
QY	3717	CCAAGAGATTCGAAACCGTTTCTTATGATGAGTGGCTGAGGATGAGGGCTTAT	3776
DB	3535	CCAAGAGATTCGAAACCGTTTCTTATGATGAGTGGCTGAGGATGAGGGCTTAT	3594
QY	3777	CGAGACAAATTTGAAGTTTGGAC	3799
DB	3595	CGAGACAAATTTGAAGTTAGTAGTC	3617

Search completed: November 22, 2004, 10:44:27
Job time : 434 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 06:14:52 ; Search time 2612 Seconds
(without alignments)
11578.897 Million cell updates/sec

Title: US-10-023-888-3

Perfect score: 5597

Sequence: 1 cggagcgcagcggcgctccg.....aaaagttaattttgaaa 5597

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 270181610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10F_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	5597	100.0	5597	9 US-09-895-072-4	Sequence 4, Appli
2	5597	100.0	5597	9 US-09-986-552-4	Sequence 4, Appli
3	5597	100.0	5597	15 US-10-023-888-3	Sequence 3, Appli
4	5597	100.0	5597	15 US-10-023-889-3	Sequence 3, Appli
5	5597	100.0	5597	15 US-10-023-890-3	Sequence 3, Appli
6	5597	100.0	5597	15 US-10-024-197-3	Sequence 3, Appli
7	5597	100.0	5597	15 US-10-023-894-3	Sequence 3, Appli
8	5597	100.0	5597	15 US-10-306-686-4	Sequence 4, Appli
9	3761	67.2	3783	9 US-09-895-072-20	Sequence 20, Appli
10	3761	67.2	3783	9 US-09-986-552-20	Sequence 20, Appli
11	3761	67.2	3783	15 US-10-306-686-20	Sequence 20, Appli
12	3485.4	62.3	3621	9 US-09-895-072-21	Sequence 21, Appli

13	3485.4	62.3	3621	9 US-09-986-552-21	Sequence 21, Appli
14	3485.4	62.3	3621	15 US-10-306-686-21	Sequence 21, Appli
15	3465.6	61.9	3600	15 US-10-023-888-1	Sequence 1, Appli
16	3465.6	61.9	3600	15 US-10-023-889-1	Sequence 1, Appli
17	3465.6	61.9	3600	15 US-10-023-890-1	Sequence 1, Appli
18	3465.6	61.9	3600	15 US-10-024-197-1	Sequence 1, Appli
19	3465.6	61.9	3600	15 US-10-023-894-1	Sequence 1, Appli
20	2820.8	50.4	5229	9 US-09-895-072-16	Sequence 16, Appli
21	2820.8	50.4	5229	9 US-09-986-552-16	Sequence 16, Appli
22	2820.8	50.4	5229	15 US-10-023-888-8	Sequence 8, Appli
23	2820.8	50.4	5229	15 US-10-023-889-8	Sequence 8, Appli
24	2820.8	50.4	5229	15 US-10-023-890-8	Sequence 8, Appli
25	2820.8	50.4	5229	15 US-10-024-197-8	Sequence 8, Appli
26	2820.8	50.4	5229	15 US-10-023-894-8	Sequence 8, Appli
27	2820.8	50.4	5229	15 US-10-306-686-16	Sequence 16, Appli
28	2613	46.7	8811	15 US-10-120-801-25	Sequence 25, Appli
29	2597.4	46.2	2603	15 US-10-094-749-753	Sequence 753, Appli
30	1977.2	35.3	6447	15 US-10-172-118-55	Sequence 59, Appli
31	1977.2	35.3	6447	16 US-10-342-887-59	Sequence 59, Appli
32	1105.2	19.7	1194	15 US-10-029-386-22697	Sequence 22697, A
33	1104.8	19.7	1962	9 US-09-864-761-10195	Sequence 10195, A
34	1103.4	19.7	1125	9 US-09-864-761-26835	Sequence 26835, A
35	724.8	12.9	728	15 US-10-139-794-170	Sequence 170, Appli
36	415.4	7.4	455	10 US-09-918-995-25442	Sequence 25442, A
37	377.2	6.7	459	10 US-09-918-995-27675	Sequence 27675, A
38	339	6.1	572	11 US-09-969-034-612	Sequence 612, Appli
39	301.4	5.4	572	15 US-10-029-386-8979	Sequence 8979, Ap
40	256.4	4.6	259	16 US-10-305-720-402	Sequence 402, Appli
41	256	4.6	337	13 US-10-079-623-97	Sequence 97, Appli
42	255	4.6	601	9 US-09-833-381-2047	Sequence 2047, Ap
43	238.2	4.3	44848	9 US-09-776-874A-42	Sequence 42, Appli
44	238.2	4.3	44848	9 US-09-988-113-42	Sequence 42, Appli
45	238.2	4.3	44848	15 US-10-341-582-42	Sequence 42, Appli

ALIGNMENTS

RESULT 1

US-09-895-072-4
; Sequence 4, Application US/09895072
; Patent No. US2002002550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCES: 210119USCCONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 5597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-895-072-4

Query Match 100.0%; Score 5597; DB 9; Length 5597;
Best local Similarity 100.0%; Pred. No. 0;
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGGAGCGGAGCGGCGCTCCGTCGCGGAGCTGCAATAGCGGCGCCCGGAGGCTGTGACC	60
Db	1	CGGAGCGGAGCGGCGCTCCGTCGCGGAGCTGCAATAGCGGCGCCCGGAGGCTGTGACC	60
Qy	61	TGGCGCGGCGGCGGCGCCCGGAGCGGCGCCCGTGAATGCGGCGCTCGTGTAGCGCGCGCGCGCGC	120
Db	61	TGGCGCGGCGGCGGCGCCCGGAGCGGCGCCCGTGAATGCGGCGCTCGTGTAGCGCGCGCGCGCGC	120
Qy	121	GGCGGCTCAGGCTCCTCGGCGGCGTGGCGGTGAAGGGGTGATGCTTCAAGCTCC	180

121 GCGCGCTCAGGCTCTCTCGGGCGTGGCGTGGCGTGAAGGGGTGATGCTGTCAAGCTCC 180
181 TGCAGAGCAAACTATACCTGCTGTCGCCACAGGTATGGCTCTACGTGTCTTCTTGG 240
181 TGCAGAGCAAACTATACCTGCTGTCGCCACAGGTATGGCTCTACGTGTCTTCTTGG 240
241 GCGTGGTGTCAACATGCTCTCGGCTTCCAGTTCGAGAGGTGGTCTCGAATGGAGCC 300
241 GCGTGGTGTCAACATGCTCTCGGCTTCCAGTTCGAGAGGTGGTCTCGAATGGAGCC 300
301 GAGATCAATCACTGTTTGTGTTTGTATCCCTATAGACAAATATTTGCTGGAAGTCTTTTC 360
301 GAGATCAATCACTGTTTGTGTTTGTATCCCTATAGACAAATATTTGCTGGAAGTCTTTTC 360
361 AGAATCGGCTTGTGTCGCCATGCGGATTTGACCTGCGGTGTTTACACCTGGTGAATGGCACAG 420
361 AGAATCGGCTTGTGTCGCCATGCGGATTTGACCTGCGGTGTTTACACCTGGTGAATGGCACAG 420
421 ATCTTGAACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGGAGGGAGCGAAG 480
421 ATCTTGAACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGGAGGGAGCGAAG 480
481 CAATGAGAGAAATCCTTGGGAAAAACACACCGAACTCTAAGAAAGTGAAGCAGT 540
481 CAATGAGAGAAATCCTTGGGAAAAACACACCGAACTCTAAGAAAGTGAAGCAGT 540
541 TAGAGTGTGCTTAACACACTGCAATTAAGTTCGCAATGCTGTGACTGGACCCAGCCCTGC 600
541 TAGAGTGTGCTTAACACACTGCAATTAAGTTCGCAATGCTGTGACTGGACCCAGCCCTGC 600
601 CAGCAACATCACCCTGAAGACCTGCTCTCTTTTATCTCTTTTCTTTTCTTTTCTTTTCT 660
601 CAGCAACATCACCCTGAAGACCTGCTCTCTTTTATCTCTTTTCTTTTCTTTTCTTTTCT 660
661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCAACAACTGCTCAAGTGTGTTTGT 720
661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCAACAACTGCTCAAGTGTGTTTGT 720
721 ACAGTACTAAGAGTGTGAAGATGCCACTGCTGACCTGCTTAAAGGAAATAGCAGACAGA 780
721 ACAGTACTAAGAGTGTGAAGATGCCACTGCTGACCTGCTTAAAGGAAATAGCAGACAGA 780
781 CAGTATGAGGGGGTACTTGACACAGATTAAGAGTCCCTGGATAGTCTAATGCAAG 840
781 CAGTATGAGGGGGTACTTGACACAGATTAAGAGTCCCTGGATAGTCTAATGCAAG 840
841 ATTTGGCTTCTCAGTGGATTTCCACCACTTCAAGGAAACAAATCAACTAAACAA 900
841 ATTTGGCTTCTCAGTGGATTTCCACCACTTCAAGGAAACAAATCAACTAAACAA 900
901 AATTGCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCCAGTG 960
901 AATTGCCAGAAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCCAGTG 960
961 TAGCGCTTCTAAAATCAATCAACCCCAAGGATTTCAAGAAATGAATGAATGAATGAATGA 1020
961 TAGCGCTTCTAAAATCAATCAACCCCAAGGATTTCAAGAAATGAATGAATGAATGAATGA 1020
1021 AGAATGACCATGATGGAAGAACTGACCAATAGCTCTGATATTTATTTATGGATC 1080
1021 AGAATGACCATGATGGAAGAACTGACCAATAGCTCTGATATTTATTTATGGATC 1080
1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGAAGATA 1140
1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGAAGATA 1140
1141 ACGAAGAACTGAGTACTCATTTGGATCTATCGAGGCGATGACCATGGGTTTCSGAATA 1200
1141 ACGAAGAACTGAGTACTCATTTGGATCTATCGAGGCGATGACCATGGGTTTCSGAATA 1200
1201 TTTTCAATGTCCAAACGGGAGATTCCTCTGCTGCTGAACCTTGAACATCTCGAGTGA 1260

1201 TTTTCATTTGTACCAACGGGCGAGATTCCATCTCTGCTGAACCTTGAACATCTCTGAGTGA 1260
1261 CAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTGTCTACTCTTATGTTTCA 1320
1261 CAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTGTCTACTCTTATGTTTCA 1320
1321 CTGCTATTGAAAGTCAATTCATTCGATCGAAGGGCTGTCCGAGAGTTTATTTACCTAA 1380
1321 CTGCTATTGAAAGTCAATTCATTCGATCGAAGGGCTGTCCGAGAGTTTATTTACCTAA 1380
1381 ATGATGATGCTCATGTTTGGAGAGGATGTCTGCCAGATGATTTTTTACAGTCACTCCAAAG 1440
1381 ATGATGATGCTCATGTTTGGAGAGGATGTCTGCCAGATGATTTTTTACAGTCACTCCAAAG 1440
1441 GCCAGAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCT 1500
1441 GCCAGAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCT 1500
1501 GGATTAAGGATGGCTATTGTGACAGGCTGTGTAATATTTGAGCTGCGGTGGGATGGTG 1560
1501 GGATTAAGGATGGCTATTGTGACAGGCTGTGTAATATTTGAGCTGCGGTGGGATGGTG 1560
1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCTGCTATATTTGAGGAGGTGGAGTACTGGGA 1620
1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCTGCTATATTTGAGGAGGTGGAGTACTGGGA 1620
1621 GTATTGGAGTGGACACCCCTGGCGAGTTTGGTGGAGGAAATAACAGTGTCTCTTACTGTA 1680
1621 GTATTGGAGTGGACACCCCTGGCGAGTTTGGTGGAGGAAATAACAGTGTCTCTTACTGTA 1680
1681 ATCAGGAGTGGCGAATTTCTGGCTCGCTGATAGTCTGTGACCAAGCATGCAATGTCT 1740
1681 ATCAGGAGTGGCGAATTTCTGGCTCGCTGATAGTCTGTGACCAAGCATGCAATGTCT 1740
1741 TGTCTGTGGGTTTGTATGCTGGCGAGTGGGCAAGATCATTTTCATGAATTTGATAAAG 1800
1741 TGTCTGTGGGTTTGTATGCTGGCGAGTGGGCAAGATCATTTTCATGAATTTGATAAAG 1800
1801 TGATCTTCTCCCAACCCAGACTCACTATATTTTCCAAAAGTGAATGCTGCTTAT 1860
1801 TGATCTTCTCCCAACCCAGACTCACTATATTTTCCAAAAGTGAATGCTGCTTAT 1860
1861 TCAGCTTTGCAAGAGTACCCAAAAGAGGAGTGGAGTGGCTATAGTGAACATCAATAA 1920
1861 TCAGCTTTGCAAGAGTACCCAAAAGAGGAGTGGAGTGGCTATAGTGAACATCAATAA 1920
1921 TTTGCACTGCTTATTTGCCAAAGTGGAAACCATCCACTCATATGACACAGTGGAA 1980
1921 TTTGCACTGCTTATTTGCCAAAGTGGAAACCATCCACTCATATGACACAGTGGAA 1980
1981 TGAATGCCCAACCAATATTTTATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040
1981 TGAATGCCCAACCAATATTTTATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040
2041 AAATCCAGATAACAGTGGAGTGGACACAGGGAGGACCAAACTGNAATTTACGGCCC 2100
2041 AAATCCAGATAACAGTGGAGTGGACACAGGGAGGACCAAACTGNAATTTACGGCCC 2100
2101 AGAAGGGTTAGCAAAATTTAGTTCCTCCATAACACTTTCTTCCAGAGGCGGAAATCTTT 2160
2101 AGAAGGGTTAGCAAAATTTAGTTCCTCCATAACACTTTCTTCCAGAGGCGGAAATCTTT 2160
2161 TTGAGGATATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGACATGATGTTTAACTCAA 2220
2161 TTGAGGATATTTCCCAAGAAACCGTTCCCGAAGTTTAAAGACATGATGTTTAACTCAA 2220
2221 CAAAGAGAGCCAGGAGAGGTGAATTTCCCTGTTAAATATTTTCACTCTCTTCCAAAAG 2280
2221 CAAAGAGAGCCAGGAGAGGTGAATTTCCCTGTTAAATATTTTCACTCTCTTCCAAAAG 2280
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340

QY 2341 AAGGATACAAATTTGTCCAAGTCAGCCTTGCCTGAGATCAATTTCTGTATGAACCTCACAGCATG 2400
DB 2341 AAGGATACAAATTTGTCCAAGTCAGCCTTGCCTGAGATCAATTTCTGTATGAACCTCACAGCATG 2400
QY 2401 CTAAATATAAATAATCAAGCTATATAATAACAGATGAACAAATATGACAGATTTGGTGGCTCCAC 2460
DB 2401 CTAAATATAAATAATCAAGCTATATAATAACAGATGAACAAATATGACAGATTTGGTGGCTCCAC 2460
QY 2461 AGGAAAAACAGGTTTCATATAAGCATCTTGCCTGAGATCAATTTCTGTATGAACCTCACAGCATG 2520
DB 2461 AGGAAAAACAGGTTTCATATAAGCATCTTGCCTGAGATCAATTTCTGTATGAACCTCACAGCATG 2520
QY 2521 AGAGTTGACATTTTCTGAGTGTGATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2580
DB 2521 AGAGTTGACATTTTCTGAGTGTGATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2580
QY 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGAACCTCACACCCCAAAACCA 2640
DB 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGAACCTCACACCCCAAAACCA 2640
QY 2641 TAGCGGGAATGTGACAAAAGAAAGCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700
DB 2641 TAGCGGGAATGTGACAAAAGAAAGCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700
QY 2701 TGACAAAAGAAAGAAATACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2760
DB 2701 TGACAAAAGAAAGAAATACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2760
QY 2761 CTGAAATACATAGGCGTTACTGAAAGTTACTTGGAAAGAAAGCTGACGATTAACAG 2820
DB 2761 CTGAAATACATAGGCGTTACTGAAAGTTACTTGGAAAGAAAGCTGACGATTAACAG 2820
QY 2821 ATAGTTACTTGGCTTTTGGCATGGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2880
DB 2821 ATAGTTACTTGGCTTTTGGCATGGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2880
QY 2881 AAGAGAGTCATTTGAAGACAAATTTGGCATCTTCACTGATGACAAAGAAAGAAAGAAAGAAAGAAAG 2940
DB 2881 AAGAGAGTCATTTGAAGACAAATTTGGCATCTTCACTGATGACAAAGAAAGAAAGAAAGAAAGAAAG 2940
QY 2941 AACTAAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATATAAATTTCTAATAGCAAGT 3000
DB 2941 AACTAAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATATAAATTTCTAATAGCAAGT 3000
QY 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
DB 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAGACGTCATTTTCAAAAGTGC 3120
DB 3061 TGCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAGACGTCATTTTCAAAAGTGC 3120
QY 3121 GCCATTTGAGGATATGAGTTTGGCTTCTCTATTTTATATCTCATGAGTGCAGTGC 3180
DB 3121 GCCATTTGAGGATATGAGTTTGGCTTCTCTATTTTATATCTCATGAGTGCAGTGC 3180
QY 3181 AGCCACTGAATATCTCAAGTCTTTGATGAGTTGATGATGATGATGATGATGATGATGATGATGATGATG 3240
DB 3181 AGCCACTGAATATCTCAAGTCTTTGATGAGTTGATGATGATGATGATGATGATGATGATGATGATGATG 3240
QY 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAA 3300
DB 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAA 3300
QY 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTCAAAATGCTTCTGCTGATATCA 3360
DB 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTCAAAATGCTTCTGCTGATATCA 3360
QY 3361 CGCAGCTAAATATTTCCACCACTGAGGATCTTACTATGATCCCAACCTGCCACCGG 3420
DB 3361 CGCAGCTAAATATTTCCACCACTGAGGATCTTACTATGATCCCAACCTGCCACCGG 3420

QY 3421 TCACATAAAGCTAGTAAACAAACCTGTAACCCAGTAACTGACAAAATCCCAAAAGCATATA 3480
DB 3421 TCACATAAAGCTAGTAAACAAACCTGTAACCCAGTAACTGACAAAATCCCAAAAGCATATA 3480
QY 3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
DB 3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540
QY 3541 TTGCTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAACCCCTAGGA 3600
DB 3541 TTGCTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAACCCCTAGGA 3600
QY 3601 AGTTTGTTCCTGGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAG 3660
DB 3601 AGTTTGTTCCTGGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAG 3660
QY 3661 CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCCAA 3720
DB 3661 CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCCAA 3720
QY 3721 GAGAGTATCGAAACCCGTTTCTTTCATATGATGAGCTGAGGAAATGGAGGCTTATCGAG 3780
DB 3721 GAGAGTATCGAAACCCGTTTCTTTCATATGATGAGCTGAGGAAATGGAGGCTTATCGAG 3780
QY 3781 ACAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATTCT 3840
DB 3781 ACAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATTCT 3840
QY 3841 CATTTTGTGCTGAGCAGTTAAATGCACTTAAGCGGAAAGATATTTCCAGAGGAGGATAC 3900
DB 3841 CATTTTGTGCTGAGCAGTTAAATGCACTTAAGCGGAAAGATATTTCCAGAGGAGGATAC 3900
QY 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCACTCTA 3960
DB 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCACTCTA 3960
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCCAGAGATGCTTTTGTGATGAT 4020
DB 3961 CCTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCCAGAGATGCTTTTGTGATGAT 4020
QY 4021 GCTTAGCAGTTTGGCCGGAAGAGAAATATCCAGTACCATGCTGCTTTTGTGGCATGAA 4080
DB 4021 GCTTAGCAGTTTGGCCGGAAGAGAAATATCCAGTACCATGCTGCTTTTGTGGCATGAA 4080
QY 4081 TATAGCCACTGACTAGGATTTATTTAAACCAACCCAGTGAACCTGCTGCTGCGACAGC 4140
DB 4081 TATAGCCACTGACTAGGATTTATTTAAACCAACCCAGTGAACCTGCTGCTGCGACAGC 4140
QY 4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCTGCTCATCTTTTATAAAA 4200
DB 4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCTGCTCATCTTTTATAAAA 4200
QY 4201 GGCTCAGTCAAGAGACAGCTGTTAATTTCCACAGCAATCATTTGACAGTAACTTTAT 4260
DB 4201 GGCTCAGTCAAGAGACAGCTGTTAATTTCCACAGCAATCATTTGACAGTAACTTTAT 4260
QY 4261 TAGGAGAGCTATGCGAGTGGAGTGTGTTAAGAGGCTCCAGTCTTTGATTTCCAA 4320
DB 4261 TAGGAGAGCTATGCGAGTGGAGTGTGTTAAGAGGCTCCAGTCTTTGATTTCCAA 4320
QY 4321 AGCCTTTGCTAAAGTTTGGCACTTTTTCATTTTCCATTTTAAAGTAGTTACTA 4380
DB 4321 AGCCTTTGCTAAAGTTTGGCACTTTTTCATTTTCCATTTTAAAGTAGTTACTA 4380
QY 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAGCAATTTGGATGCTTAAACCTTTTAA 4440
DB 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAGCAATTTGGATGCTTAAACCTTTTAA 4440
QY 4441 TAGATGTTATTTAAATATGACAGCAATATCACTCTTTATTGACAAATACCTTAAATATGAG 4500
DB 4441 TAGATGTTATTTAAATATGACAGCAATATCACTCTTTATTGACAAATACCTTAAATATGAG 4500
QY 4501 TTTTATTAATTTAAGACTGTAATGGTCTTAAACCACTACTACTGAGAGCTCAATG 4560

QY 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTTCTCAGTTGTTGTTTG 720
Db 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTTCTCAGTTGTTGTTTG 720
QY 721 ACAGTACTAAGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA 780
Db 721 ACAGTACTAAGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA 780
QY 781 CAGTATGAGGGGGTACTTGCACACAGATAAAGAGTCCCTGGATTAGTGTCTATGCAAG 840
Db 781 CAGTATGAGGGGGTACTTGCACACAGATAAAGAGTCCCTGGATTAGTGTCTATGCAAG 840
QY 841 ATTTGGCTTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCACTAAACAA 900
Db 841 ATTTGGCTTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCACTAAACAA 900
QY 901 AATTGCCAGAAATCTTCTCTTAAAGTCAAACCTGTTGCAGTTGTATTCAGAGGCCAGTG 960
Db 901 AATTGCCAGAAATCTTCTCTTAAAGTCAAACCTGTTGCAGTTGTATTCAGAGGCCAGTG 960
QY 961 TAGCGCTTTCTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATAAGCAAACTAAGA 1020
Db 961 TAGCGCTTTCTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATAAGCAAACTAAGA 1020
QY 1021 AGACATGACCATGATGGAAGAAAGTGAACAATAAGTCTCTGCAATATTTATATGGGATC 1080
Db 1021 AGACATGACCATGATGGAAGAAAGTGAACAATAAGTCTCTGCAATATTTATATGGGATC 1080
QY 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCCAGTCGTTTGAAGATA 1140
Db 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCCAGTCGTTTGAAGATA 1140
QY 1141 ACAGAGAACTGAGTACTCATTTGCGATCTATCGAGGSCATGCACCATTGGTTTCGGAATA 1200
Db 1141 ACAGAGAACTGAGTACTCATTTGCGATCTATCGAGGSCATGCACCATTGGTTTCGGAATA 1200
QY 1201 TTTTCATGTCACCAACGGGAGATTCATCTGCTGAGGCTGAACTTGAACATCTCGAGTGA 1260
Db 1201 TTTTCATGTCACCAACGGGAGATTCATCTGCTGAGGCTGAACTTGAACATCTCGAGTGA 1260
QY 1261 CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTCA 1320
Db 1261 CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTCA 1320
QY 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTGCCAGAGTTTATTTACCTAA 1380
Db 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTGCCAGAGTTTATTTACCTAA 1380
QY 1381 ATGATGATGTATGTTGGGAGGATGTCGCGCAGATGATTTTACAGTCACCTCCAAAG 1440
Db 1381 ATGATGATGTATGTTGGGAGGATGTCGCGCAGATGATTTTACAGTCACCTCCAAAG 1440
QY 1441 GCCAGAGGTTTATTGACATGGCTGTGCCAACTGTGCGGAGGCTGCCAGGTTTCT 1500
Db 1441 GCCAGAGGTTTATTGACATGGCTGTGCCAACTGTGCGGAGGCTGCCAGGTTTCT 1500
QY 1501 GGATTAAGGATGGTATTGTGCAAGGCTTGTATTAATTCAGCTCGATGGATGGTG 1560
Db 1501 GGATTAAGGATGGTATTGTGCAAGGCTTGTATTAATTCAGCTCGATGGATGGTG 1560
QY 1561 GGGATTGCTCTGGAACACAGTGGAGGAGTCCCTATTTGAGGAGGTTGAGGTTACTGGGA 1620
Db 1561 GGGATTGCTCTGGAACACAGTGGAGGAGTCCCTATTTGAGGAGGTTGAGGTTACTGGGA 1620
QY 1621 GTATTGGATGGGACACCCCTGGCAGTTTGTGGAGGAATAAAGTGTCTCTTACTGTA 1680
Db 1621 GTATTGGATGGGACACCCCTGGCAGTTTGTGGAGGAATAAAGTGTCTCTTACTGTA 1680
QY 1681 ATCAGGGATGTCGAAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGCT 1740
Db 1681 ATCAGGGATGTCGAAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGCT 1740
QY 1741 TGTCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCATTTTTCATGAATTTGATTAAG 1800

Db 1741 TGTCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCATTTTCATGAATTTGATTAAG 1800
QY 1801 TGAATCTTCTCCCAACACAGACTCATATATTTATTTCCAAAAGTGAATGCTGCTTATT 1860
Db 1801 TGAATCTTCTCCCAACACAGACTCATATATTTATTTCCAAAAGTGAATGCTGCTTATT 1860
QY 1861 TCAGCTTTGAGAGTAGCCAAAAGAGGAGTTGAAGTGCCTTATAGTGAATCCAAATA 1920
Db 1861 TCAGCTTTGAGAGTAGCCAAAAGAGGAGTTGAAGTGCCTTATAGTGAATCCAAATA 1920
QY 1921 TTCGACATGCTTTCTATTGCCAAACAGTGGAAAACCATCCACCTCATATAATGCAAGTGA 1980
Db 1921 TTCGACATGCTTTCTATTGCCAAACAGTGGAAAACCATCCACCTCATATAATGCAAGTGA 1980
QY 1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTCCAAAATCAAAACGATGAAGTGTCA 2040
Db 1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTCCAAAATCAAAACGATGAAGTGTCA 2040
QY 2041 AAATGACAGATAACAGTGGAGGTGGACACAAGGAGGACCAAACTGAAATTTACGGCC 2100
Db 2041 AAATGACAGATAACAGTGGAGGTGGACACAAGGAGGACCAAACTGAAATTTACGGCC 2100
QY 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCATTAACATCTTCCAGAGGCGGAAATCCTTT 2160
Db 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCATTAACATCTTCCAGAGGCGGAAATCCTTT 2160
QY 2161 TTGAGGATATTTCCAAAAGAAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220
Db 2161 TTGAGGATATTTCCAAAAGAAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220
QY 2221 CAAGGAGAGCCAGGAGGAGTGAATAATCCCTGGTAAATTTTCACTCTTCCAAAAG 2280
Db 2221 CAAGGAGAGCCAGGAGGAGTGAATAATCCCTGGTAAATTTTCACTCTTCCAAAAG 2280
QY 2281 AGCCAGTGTGATCTCAATACCTGGATTTCCAACTGGAACATGAGGACATCCTTGA 2340
Db 2281 AGCCAGTGTGATCTCAATACCTGGATTTCCAACTGGAACATGAGGACATCCTTGA 2340
QY 2341 AAGGATACAAATTTGTCCTCAGAGTCAATTTCTGATGAATTTCTGATGAATTTGATG 2400
Db 2341 AAGGATACAAATTTGTCCTCAGAGTCAATTTCTGATGAATTTCTGATGAATTTGATG 2400
QY 2401 CTAAATTAATAATTAAGCTTAATAACAGATGAACATGAGTGGTGGTGGTGGTGGTGG 2460
Db 2401 CTAAATTAATAATTAAGCTTAATAACAGATGAACATGAGTGGTGGTGGTGGTGGTGG 2460
QY 2461 AGGAAAACAGGTTTCAATAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTC 2520
Db 2461 AGGAAAACAGGTTTCAATAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTC 2520
QY 2521 AGAGTTGATCTTTTCTGAGTGTGTAAGTGAATGGTCAATGAGGTTGAGGTTGAGGTT 2580
Db 2521 AGAGTTGATCTTTTCTGAGTGTGTAAGTGAATGGTCAATGAGGTTGAGGTTGAGGTT 2580
QY 2581 CACCCCTGACCTTGGAGACCAAGATTTAGAGTGAATTTAGAGTGGAACTCACACCAAAACCA 2640
Db 2581 CACCCCTGACCTTGGAGACCAAGATTTAGAGTGGAACTCACACCAAAACCA 2640
QY 2641 TAGCGGAAATGTGCAAAAGAAAAGCCCTCTCTGATTTCTGATTTCTGATTTCTGATTT 2700
Db 2641 TAGCGGAAATGTGCAAAAGAAAAGCCCTCTCTGATTTCTGATTTCTGATTTCTGATTT 2700
QY 2701 TGAACAAAGAAAAGAAATCAAGGGAAGAAAAGAGACAGATGAGATGGAGGAAATG 2760
Db 2701 TGAACAAAGAAAAGAAATCAAGGGAAGAAAAGAGACAGATGAGATGGAGGAAATG 2760
QY 2761 CTGAAATCAGATAGGCTTACTGAGTGTACTTCTGAAAGAAAGTGGAGCAATACACAG 2820
Db 2761 CTGAAATCAGATAGGCTTACTGAGTGTACTTCTGAAAGAAAGTGGAGCAATACACAG 2820
QY 2821 ATAGTTACTTGGCTTTTGGCCATGGGAGAAAAGATTTTCCAGATTTCTCTGAGG 2880

Db 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAAGATATTTCCAAAGATCTTCTCGACG 2880
Qy 2881 AAGAAGAGTCTATTGAAGACACAAATGGCATACTTCTCACTGATAGCAAAAAATCTCTGGAGGC 2940
Db 2881 AAGAAGAGTCTATTGAAGACACAAATGGCATACTTCTCACTGATAGCAAAAAATCTCTGGAGGC 2940
Qy 2941 AACTTAAAGATACATTTGCGAGATTTCCCTCAGATATGTAATAAATTTCTTAAATAGCAAGT 3000
Db 2941 AACTTAAAGATACATTTGCGAGATTTCCCTCAGATATGTAATAAATTTCTTAAATAGCAAGT 3000
Qy 3001 TTGATTCACATCGCGAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
Db 3001 TTGATTCACATCGCGAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060
Qy 3061 TGCAGAACTCCAGATATATGTTCCCTGAGAAATTTGACAGCTCATTTTCAAAAGTGC 3120
Db 3061 TGCAGAACTCCAGATATATGTTCCCTGAGAAATTTGACAGCTCATTTTCAAAAGTGC 3120
Qy 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGCAGTGC 3180
Db 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGCAGTGC 3180
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTCTGT 3240
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTCTGT 3240
Qy 3241 CTGACAGAGAAATCGGAACATCTGCTACCAAGATTCAGAACTGCCGTTAAGTTTCAGG 3300
Db 3241 CTGACAGAGAAATCGGAACATCTGCTACCAAGATTCAGAACTGCCGTTAAGTTTCAGG 3300
Qy 3301 ATTTGACAGGCTCGGAAACATCTGATTAATTAATTTCTCAAAAATGCTTCCCTGCTGATATCA 3360
Db 3301 ATTTGACAGGCTCGGAAACATCTGATTAATTAATTTCTCAAAAATGCTTCCCTGCTGATATCA 3360
Qy 3361 CGCAGCTAAATTAATTTCCACCAATCTCGGAATCCTACTATGATCCCAACCTGCCACCGG 3420
Db 3361 CGCAGCTAAATTAATTTCCACCAATCTCGGAATCCTACTATGATCCCAACCTGCCACCGG 3420
Qy 3421 TCATCTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCATATA 3480
Db 3421 TCATCTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCATATA 3480
Qy 3481 AGGACAAAAAACAATATAGTTTGAATCATGAGGAGAGAAATCGCTTTTAAATGA 3540
Db 3481 AGGACAAAAAACAATATAGTTTGAATCATGAGGAGAGAAATCGCTTTTAAATGA 3540
Qy 3541 TTCGTACCAAGTTCCTCATGTTGTCGAGTTGGATGACATGAAGAAAAACCTTAGGA 3600
Db 3541 TTCGTACCAAGTTCCTCATGTTGTCGAGTTGGATGACATGAAGAAAAACCTTAGGA 3600
Qy 3601 AGTTTGTGCTGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660
Db 3601 AGTTTGTGCTGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660
Qy 3661 CTGTTCTCAGGACTTCTATGAATCCATGTTCCCTCATCTTCCCAATTTGAATCGCAA 3720
Db 3661 CTGTTCTCAGGACTTCTATGAATCCATGTTCCCTCATCTTCCCAATTTGAATCGCAA 3720
Qy 3721 GAGAGTATCGAAACCGTTTCTTCTCATATGATGAGTGCAGGAATGGAGGCTTATCGAG 3780
Db 3721 GAGAGTATCGAAACCGTTTCTTCTCATATGATGAGTGCAGGAATGGAGGCTTATCGAG 3780
Qy 3781 ACAAAATGAGTTTGGACCAATTTGTHACTAGCAACATTAAGTATGTTTACTATATCT 3840
Db 3781 ACAAAATGAGTTTGGACCAATTTGTHACTAGCAACATTAAGTATGTTTACTATATCT 3840
Qy 3841 CATTTTTGCTGAGCAGTTTAAATGACCTTAAGCGGAGATATTTCCAGAGAGGAGTAC 3900
Db 3841 CATTTTTGCTGAGCAGTTTAAATGACCTTAAGCGGAGATATTTCCAGAGAGGAGTAC 3900
Qy 3901 ACAAGAGAGTGTCTCCCAATCGAATCAGAGATATGAGAGATCTTCAATTTGAAACCATCTA 3960
Db 3901 ACAAGAGAGTGTCTCCCAATCGAATCAGAGATATGAGAGATCTTCAATTTGAAACCATCTA 3960

Qy 3961 CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGAGATGCTCTTTGTGATGTGAT 4020
Db 3961 CCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGAGATGCTCTTTGTGATGTGAT 4020
Qy 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTTTGTGCGCATGAA 4080
Db 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTTTGTGCGCATGAA 4080
Qy 4081 TATAGCCCACTGACTAGGAAATTTTAAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140
Db 4081 TATAGCCCACTGACTAGGAAATTTTAAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140
Qy 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGCTCATCTCTTTTATAAAA 4200
Db 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGCTCATCTCTTTTATAAAA 4200
Qy 4201 GGCTCACTGACAAGAGACAGCTGTTTAAATTTCCACAGCAATCATTCGACAGATACTTTAT 4260
Db 4201 GGCTCACTGACAAGAGACAGCTGTTTAAATTTCCACAGCAATCATTCGACAGATACTTTAT 4260
Qy 4261 TAGGAGAACCTTATGCCAGCTGGGAGTGAATTTGCTAAGAGGCTCCAGTCTTTTGCATTCGAA 4320
Db 4261 TAGGAGAACCTTATGCCAGCTGGGAGTGAATTTGCTAAGAGGCTCCAGTCTTTTGCATTCGAA 4320
Qy 4321 AGCCCTTTTGTCTAAAGTTTTGCACTTTTTTTTTTTTCAITTTCCCAITTTTTTAAAGTAGTACTA 4380
Db 4321 AGCCCTTTTGTCTAAAGTTTTGCACTTTTTTTTTTTTCAITTTCCCAITTTTTTAAAGTAGTACTA 4380
Qy 4381 AGTTAACTAGTATTTCTTGTCTTCTGAGTATTAACGAATTTGGGATGCTTAAACCTTATTTTA 4440
Db 4381 AGTTAACTAGTATTTCTTGTCTTCTGAGTATTAACGAATTTGGGATGCTTAAACCTTATTTTA 4440
Qy 4441 TAGATGTTTATTTAAATAATGAGCAATATCACTCTTATTGACCAATACCTAAATTAATGAG 4500
Db 4441 TAGATGTTTATTTAAATAATGAGCAATATCACTCTTATTGACCAATACCTAAATTAATGAG 4500
Qy 4501 TTTTATTAATATTTAAGACTGTAAATGCTTTAAACCACTAACTACTGAAAGAGTCAATG 4560
Db 4501 TTTTATTAATATTTAAGACTGTAAATGCTTTAAACCACTAACTACTGAAAGAGTCAATG 4560
Qy 4561 ATTGACATCTGAAATGCTTTTGAATTTATTGATCTTCCAGCCCTTAAAGATGCTATGATTTCA 4620
Db 4561 ATTGACATCTGAAATGCTTTTGAATTTATTGATCTTCCAGCCCTTAAAGATGCTATGATTTCA 4620
Qy 4621 CGTGCAGGTCTAAATTTCAACAGGCTTAGATTAAGTACTACTTACCAGATGTAATTAATGTTT 4680
Db 4621 CGTGCAGGTCTAAATTTCAACAGGCTTAGATTAAGTACTACTTACCAGATGTAATTAATGTTT 4680
Qy 4681 TGGAAATGTACATATTTCAACAGAGTCCCTCATTTTGAAGATGAGTATGCTGATGCGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGTCCCTCATTTTGAAGATGAGTATGCTGATGCGCA 4740
Qy 4741 CTGGCACAATTACAGTGGTGTCTTTTAAATACTCATTTGCTATATTTCCAGTAGTATCTCT 4800
Db 4741 CTGGCACAATTACAGTGGTGTCTTTTAAATACTCATTTGCTATATTTCCAGTAGTATCTCT 4800
Qy 4801 CTCAGTTGGTTTTTGAATAGAACAGAGGCGCAAACTTTCTTTGTAAGAGGCTGGTGTAGT 4860
Db 4801 CTCAGTTGGTTTTTGAATAGAACAGAGGCGCAAACTTTCTTTGTAAGAGGCTGGTGTAGT 4860
Qy 4861 AAATTAATTTGAGGCGCACCTGCTCTTTGTCTATACATTTCTTCTTGTGTTTGTAGTTGT 4920
Db 4861 AAATTAATTTGAGGCGCACCTGCTCTTTGTCTATACATTTCTTCTTGTGTTTGTAGTTGT 4920
Qy 4921 TTTTCTTCAACCAACCTCTTAAATAATGTAATAAACAATGTTTAGCTTGAGCTGTACAAAA 4980
Db 4921 TTTTCTTCAACCAACCTCTTAAATAATGTAATAAACAATGTTTAGCTTGAGCTGTACAAAA 4980
Qy 4981 ACTGCGCCACCGAGATGTCAGGCTCATTTTGCACATCTCTGAGAAATATTTT 5040
Db 4981 ACTGCGCCACCGAGATGTCAGGCTCATTTTGCAGCCATCTTTGCAATCTCTGAGAAATATTTT 5040

QY 1261 CAATAGTAACACACAGAGATGTTTTGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320
DB 1261 CAATAGTAACACACAGAGATGTTTTGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320
QY 1321 CTGCTATTGAAAGTCACATTCGCAATCGAAGGGCTGTCCAGAAAGTTTATTACCTAA 1380
DB 1321 CTGCTATTGAAAGTCACATTCGCAATCGAAGGGCTGTCCAGAAAGTTTATTACCTAA 1380
QY 1381 ATGATGATGTCATGTTGGGAAGGATGTCCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
DB 1381 ATGATGATGTCATGTTGGGAAGGATGTCCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
QY 1441 GCCAGAGGTTTATTGACATGGGCTGTGCCAACTGTGCCAGGGCTGCCAGGTTTCCT 1500
DB 1441 GCCAGAGGTTTATTGACATGGGCTGTGCCAACTGTGCCAGGGCTGCCAGGTTTCCT 1500
QY 1501 GGATTAAGGATGGCTATTGTGACAAAGCTTGTAAATTTGAGCTCGATTTGGGATGGTG 1560
DB 1501 GGATTAAGGATGGCTATTGTGACAAAGCTTGTAAATTTGAGCTCGATTTGGGATGGTG 1560
QY 1561 GGGATTTGCTCTGGAAACAGTGGAGGAGTGCCTATATTGAGGAGGTGGAGTACTGGGA 1620
DB 1561 GGGATTTGCTCTGGAAACAGTGGAGGAGTGCCTATATTGAGGAGGTGGAGTACTGGGA 1620
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680
DB 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680
QY 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740
DB 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740
QY 1741 TGTCTGTGGGTTGATCTGGCAGCTGTGGCAGAGATCAATTTTCAATCAATTTGATATAAG 1800
DB 1741 TGTCTGTGGGTTGATCTGGCAGCTGTGGCAGAGATCAATTTTCAATCAATTTGATATAAG 1800
QY 1801 TGATCTTTCTCCAAACAGACTGATATATATTATTCGAAAGGTAATGCTGCTTATT 1860
DB 1801 TGATCTTTCTCCAAACAGACTGATATATATTATTCGAAAGGTAATGCTGCTTATT 1860
QY 1861 TCAGCTTTGCAAGATGACCAAAAGAGAGTTGAGGTGCCTATAGTGACATCCAAATAA 1920
DB 1861 TCAGCTTTGCAAGATGACCAAAAGAGAGTTGAGGTGCCTATAGTGACATCCAAATAA 1920
QY 1921 TTCGACATGCTTCTATTGCCAAAGTGGAAAAACATCCACCTCATATGTCACAGTGGAA 1980
DB 1921 TTCGACATGCTTCTATTGCCAAAGTGGAAAAACATCCACCTCATATGTCACAGTGGAA 1980
QY 1981 TGAATGCCACACATATATTTAATCTCAGCTTTCAAAATACAAACGATGAAGTTCA 2040
DB 1981 TGAATGCCACACATATATTTAATCTCAGCTTTCAAAATACAAACGATGAAGTTCA 2040
QY 2041 AAATCAGATTAACAGTGGAGTGGACACAGGGAGGGACCAAAACTGAAATCTACGGCC 2100
DB 2041 AAATCAGATTAACAGTGGAGTGGACACAGGGAGGGACCAAAACTGAAATCTACGGCC 2100
QY 2101 AGAAGGTTACGAAATTTAGTTAGTCCATTAACATCTTCCAGAGCGGAATCCTTT 2160
DB 2101 AGAAGGTTACGAAATTTAGTTAGTCCATTAACATCTTCCAGAGCGGAATCCTTT 2160
QY 2161 TTGAGGATATCCCAAGAAAAACGCTTCCCGAAGTTTAAGAGACATGATTTAACTCAA 2220
DB 2161 TTGAGGATATCCCAAGAAAAACGCTTCCCGAAGTTTAAGAGACATGATTTAACTCAA 2220
QY 2221 CAAGGAGAGCCAGGAAGGTGAAATTTCCCTGGTAAATATTTCTCTCCCTCCAAAG 2280
DB 2221 CAAGGAGAGCCAGGAAGGTGAAATTTCCCTGGTAAATATTTCTCTCCCTCCAAAG 2280
QY 2281 ACGCCAGTGTAGTCTCAATCTTGGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340
DB 2281 ACGCCAGTGTAGTCTCAATCTTGGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340

QY 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAATTTCTGATGAACCTACAGCATG 2400
DB 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAATTTCTGATGAACCTACAGCATG 2400
QY 2401 CTAAATAAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGTGTGCTCCAC 2460
DB 2401 CTAAATAAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGTGTGCTCCAC 2460
QY 2461 AGGAAAAACAGGTTTCATATAAGCATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
DB 2461 AGGAAAAACAGGTTTCATATAAGCATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520
QY 2521 AGAGTTGACTTTTCTCGAGTGTGTAAGTAATGTCATGATGTCACAGGGTCAGAAATC 2580
DB 2521 AGAGTTGACTTTTCTCGAGTGTGTAAGTAATGTCATGATGTCACAGGGTCAGAAATC 2580
QY 2581 CACCCCTGGACTTGGAGACCCAGCAAGATTTAGAGTGGAACTCACACCCCAAAAACCA 2640
DB 2581 CACCCCTGGACTTGGAGACCCAGCAAGATTTAGAGTGGAACTCACACCCCAAAAACCA 2640
QY 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTCAATTTGCTTCCACTGGAAAGCCAGA 2700
DB 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTCAATTTGCTTCCACTGGAAAGCCAGA 2700
QY 2701 TGACAAAAGAAAAGAAAATCACAGGAAAGAAAAGAGAAACAGTAGATGAGGAAATG 2760
DB 2701 TGACAAAAGAAAAGAAAATCACAGGAAAGAAAAGAGAAACAGTAGATGAGGAAATG 2760
QY 2761 CTGAAAATCACATAGGCGTTACTGAAAGTTTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820
DB 2761 CTGAAAATCACATAGGCGTTACTGAAAGTTTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820
QY 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAAGTATTTCCAGATCTTCTCGAG 2880
DB 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAAGTATTTCCAGATCTTCTCGAG 2880
QY 2881 AAGAAGAGTCATTGAAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC 2940
DB 2881 AAGAAGAGTCATTGAAAGACACAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC 2940
QY 2941 AACTAAAAAGATACATTTGACAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
DB 2941 AACTAAAAAGATACATTTGACAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
QY 3001 TTGGATTCATATCGGAAAGTCCCTGCTCATGCTCATGATTTGACCGGATTTGTTA 3060
DB 3001 TTGGATTCATATCGGAAAGTCCCTGCTCATGCTCATGATTTGACCGGATTTGTTA 3060
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGCGTCAATTCACAAAGTGC 3120
DB 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGCGTCAATTCACAAAGTGC 3120
QY 3121 GGCATTTGAGATATGAGTTTGGCTTCTCTTAATTTTATCTCATGATGAGTGCAGTGC 3180
DB 3121 GGCATTTGAGATATGAGTTTGGCTTCTCTTAATTTTATCTCATGATGAGTGCAGTGC 3180
QY 3181 AGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240
DB 3181 AGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240
QY 3241 CTGACAGAAATCCGACACATCGGCTACAGAAATTCAGAACTGCGGTAAAGTTGCGAG 3300
DB 3241 CTGACAGAAATCCGACACATCGGCTACAGAAATTCAGAACTGCGGTAAAGTTGCGAG 3300
QY 3301 ATTTGACAGCTGAGAAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360
DB 3301 ATTTGACAGCTGAGAAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360
QY 3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420
DB 3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420
QY 3421 TCACTAAAAGTCTAGTAACAAACTGTAAACAGTAATGACAAAATCCCAAAAGCATATA 3480

[illegible]

Db	2881	AAGAAGAGTCATTGAAGACACAAATGGCATCTTCTACTGATAGCAAAAATPACTGGGAGGC	2944
Qy	2941	AACTAAAGAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAAAATCTAAATAGCAAGT	3000
Db	2941	AACTAAAGAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAAAATCTAAATAGCAAGT	3000
Qy	3001	TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCTCACAATGATGACCGGATTTGTTA	3060
Db	3001	TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCTCACAATGATGACCGGATTTGTTA	3060
Qy	3061	TGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGCAAGAGCGTCATTTCCAAAAAGTGC	3120
Db	3061	TGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGCAAGAGCGTCATTTCCAAAAAGTGC	3120
Qy	3121	GCCATTCTGAGGATATGCAAGTTCGCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC	3180
Db	3121	GCCATTCTGAGGATATGCAAGTTCGCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC	3180
Qy	3181	AGGCACCTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT	3240
Db	3181	AGGCACCTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT	3240
Qy	3241	CTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCCGTTAAAGTTGCAGG	3300
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCCGTTAAAGTTGCAGG	3300
Qy	3301	ATTGTGACAGCTCGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCCTGCTGATATCA	3360
Db	3301	ATTGTGACAGCTCGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCCTGCTGATATCA	3360
Qy	3361	CGCAGCTTAAATAATTTCCCAACCTCAGAAATCCTACTATGATCCCAACCTGCCACCGG	3420
Db	3361	CGCAGCTTAAATAATTTCCCAACCTCAGAAATCCTACTATGATCCCAACCTGCCACCGG	3420
Qy	3421	TCACTAAAAGCTAGTAAACAAACGTGTAACACAGTAACTGACAAATCCACAAAGCATATA	3480
Db	3421	TCACTAAAAGCTAGTAAACAAACGTGTAACACAGTAACTGACAAATCCACAAAGCATATA	3480
Qy	3481	AGACAAAAACAATATAGGTTTCAATCATGCGAAGAAAGAAATCGCTTTTAAATATGA	3540
Db	3481	AGACAAAAACAATATAGGTTTCAATCATGCGAAGAAAGAAATCGCTTTTAAATATGA	3540
Qy	3541	TTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAGAAAACCCCTAGGA	3600
Db	3541	TTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAGAAAACCCCTAGGA	3600
Qy	3601	AGTTTGTTCCTGGAATGACAAACATTTGACCAACATCATAAAGATGCTCAGACAGTGAAGG	3660
Db	3601	AGTTTGTTCCTGGAATGACAAACATTTGACCAACATCATAAAGATGCTCAGACAGTGAAGG	3660
Qy	3661	CTGTCTCTCGGGACTTCTATGAATCCATGTTCCCAATCCTTCCCAATTTGAACTGCCAA	3720
Db	3661	CTGTCTCTCGGGACTTCTATGAATCCATGTTCCCAATCCTTCCCAATTTGAACTGCCAA	3720
Qy	3721	GAGAGTACGAACCGTTTCCCTCATATGATCATGAGCTGCAAGGAATGAGGGCTTATCGAG	3780
Db	3721	GAGAGTACGAACCGTTTCCCTCATATGATCATGAGCTGCAAGGAATGAGGGCTTATCGAG	3780
Qy	3781	ACAAATTTGAAGTTTTTGGACCCATGTGTACTAGCAACATTTGATTATGTTTACTATATTCT	3840
Db	3781	ACAAATTTGAAGTTTTTGGACCCATGTGTACTAGCAACATTTGATTATGTTTACTATATTCT	3840
Qy	3841	CATTTTTTGTGACAGTTAAATTTGCACTTAAAGCGGAGATATTTCCCGAGAGGAGATAC	3900
Db	3841	CATTTTTTGTGACAGTTAAATTTGCACTTAAAGCGGAGATATTTCCCGAGAGGAGATAC	3900
Qy	3901	ACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAGAATCTTCATTGTAAGAACCATCTA	3960
Db	3901	ACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAGAATCTTCATTGTAAGAACCATCTA	3960
Qy	3961	CCTCAGCATTTACTGACATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTGAT	4020
Db	3961	CCTCAGCATTTACTGACATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTGAT	4020

QY	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
DB	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
QY	4081	TATAGCCCACTGACTAGGAATATTTTAAACCAACCACCTGAAACCTCTGTGTGCGAGCAGC	4140
DB	4081	TATAGCCCACTGACTAGGAATATTTTAAACCAACCACCTGAAACCTGTGTGTGCGAGCAGC	4140
QY	4141	TCTGAACCTGATTTTACTTTTTTAAAGAAATTTGCTCATGAGCCGTGCATCCTTTTTATAAAAA	4200
DB	4141	TCTGAACCTGATTTTACTTTTTTAAAGAAATTTGCTCATGAGCCGTGCATCCTTTTTATAAAAA	4200
QY	4201	GGCTCACTGACAAGACAGACGTGTTAATTTCCCAACAGCAATCATTTGCAGACTCACTTTTAT	4260
DB	4201	GGCTCACTGACAAGACAGACGTGTTAATTTCCCAACAGCAATCATTTGCAGACTCACTTTTAT	4260
QY	4261	TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTPAAGAGGCTCCAGTCTTTGTCATTCCAA	4320
DB	4261	TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTPAAGAGGCTCCAGTCTTTGTCATTCCAA	4320
QY	4321	AGCCTTTTGCTTAAGTTTTCGACATTTTTTTTTTTCATTTCCCATTTTAAAGTAGTAGTACTA	4380
DB	4321	AGCCTTTTGCTTAAGTTTTCGACATTTTTTTTTTTCATTTCCCATTTTAAAGTAGTAGTACTA	4380
QY	4381	AGTTAACTAGTTATCTTGTCTTCGAGTATAACGAAATGGGATGCTCTAAACCTATTTTTA	4440
DB	4381	AGTTAACTAGTTATCTTGTCTTCGAGTATAACGAAATGGGATGCTCTAAACCTATTTTTA	4440
QY	4441	TAGATGTTAATTTAAATAATGCAGCAATATCACCTCTTATTTGACAAATACCTAAATATTAG	4500
DB	4441	TAGATGTTAATTTAAATAATGCAGCAATATCACCTCTTATTTGACAAATACCTAAATATTAG	4500
QY	4501	TTTTTATTAATTTTAAAGACTGTAATGGTCTTTAAACCACCTAACTACTGGAAGAGCTCAATG	4560
DB	4501	TTTTTATTAATTTTAAAGACTGTAATGGTCTTTAAACCACCTAACTACTGGAAGAGCTCAATG	4560
QY	4561	ATTGACATCTGAATGCTTTGTATTAATTAATGACCTTTCAGCCCTAAGANTGCTATGATTCA	4620
DB	4561	ATTGACATCTGAATGCTTTGTATTAATTAATGACCTTTCAGCCCTAAGANTGCTATGATTCA	4620
QY	4621	CGTGCAGGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACAGATGTAATATATGTTT	4680
DB	4621	CGTGCAGGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACAGATGTAATATATGTTT	4680
QY	4681	TGGAATATGTACATATTCAACAGAGTGCCTCATTTTAGAANAAGTAGTAGTCTGATGGCA	4740
DB	4681	TGGAATATGTACATATTCAACAGAGTGCCTCATTTTAGAANAAGTAGTAGTCTGATGGCA	4740
QY	4741	CTGGCACAATACAGTGGTGCTCTGTTTAAATACTCATTGGTATATTCAGTAGCTATCTCT	4800
DB	4741	CTGGCACAATACAGTGGTGCTCTGTTTAAATACTCATTGGTATATTCAGTAGCTATCTCT	4800
QY	4801	CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT	4860
DB	4801	CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT	4860
QY	4861	AAATTAATGAGGCCACCTGTGCTTTGTGCATACATCTCTTCTGCTCTGTTTAGTTTGT	4920
DB	4861	AAATTAATGAGGCCACCTGTGCTTTGTGCATACATCTCTTCTGCTCTGTTTAGTTTGT	4920
QY	4921	TTTTTTTTCAACAAACCTCTAAATAATGTAATAAACCATGTTTAGCTTGCAGCTGTACAAAA	4980
DB	4921	TTTTTTTTCAACAAACCTCTAAATAATGTAATAAACCATGTTTAGCTTGCAGCTGTACAAAA	4980
QY	4981	ACTGCCACACAGCCAGATGTGACCCCTCAGGCGCATCATTTGCCAATCACTCAGAAATTATTT	5040
DB	4981	ACTGCCACACAGCCAGATGTGACCCCTCAGGCGCATCATTTGCCAATCACTCAGAAATTATTT	5040
QY	5041	TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTTGCACAGCTGGAG	5100
DB	5041	TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTTGCACAGCTGGAG	5100

RESULT 5

RESULT 5
US-10-023-890-3

US-10-023-890-3
Communication III / 1 000000

; Sequence 3, Application US/10022

; Publication No. US20

; GENERAL INFORMATION:

APPLICANT: CANFIELD, William

; TITLE OF INVENTION: METHOD OF PRODUCING...

: TITLE OF INVENTION: MAMMALIAN

; TITLE OF INVENTION: MATHEMATIAN CELLS
 : FILE REFERENCE: 203510US77

```

; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: 115/1

```

; CURRENT APPLICATION NUMBER

; CURRENT FILING DATE: 2001-12-

; NUMBER OF S

; SOFTWARE: Pat

; SEQ ID NO 3

; LENGTH: 5597

: TYPE: DNA

Query Match 100.0% Score 5597 DB 15 Length 5597:

Query Match

Best Local Similarity 100.0%; pred. No. 0;

1 CCGAGCCGAGCGGGCGTCGTCGCCCGGAGCTGCAATGAGCGGGCGCCGGAGGCTGTGACC 60

QY I CGGAGCCGAGCGGGGTCGCTGGAGCTGCAATGAGCGGCGCCGGAGGCTGTGACC 80

61 TGGGGGGGGGGCGGCGGACCGGGGGCCCTGAATGGCGGCTCGCTAGGGGGGGGGCGGC 120

QY
6I
12

121 GGCGGCTCAGGCTCCTCGGGCGTGGCGGTGAAGGGTGATGCTGTTCAGCTCC 180

QY IZI GGCCTCAGGCTCTCCGGGCGTGGCGTGACGGTGAATGTGTTCTTAAGCTCC TGG

181 TGCAGAGACAAACCTATACCTGCCCTGTCCACAGGTATGGGCTCTACGTGTGCTTCTTG 240

181 TGAGAGACAAACCTTAATACCTGCTGTCCACAGGTATGGGCTCTACGTTGCTTCTTGG 240
241 GCGTGTGTCACCATCGTCTCCGCTTCCAGTTCGAGAGGTGTTCTTGAATGGAGCC 300
241 GCGTGTGTCACCATCGTCTCCGCTTCCAGTTCGAGAGGTGTTCTTGAATGGAGCC 300
301 GAGATCAATACCATGTTTTGTTGATTCCTATAGAGACATATTCGTCGAAGTCCCTTC 360
301 GAGATCAATACCATGTTTTGTTGATTCCTATAGAGACATATTCGTCGAAGTCCCTTC 360
361 AGAATCGGCTTTGTCGCCCATCGCAATGACGTTGTTTACACCTCGGTTGAATGCCACAG 420
361 AGAATCGGCTTTGTCGCCCATCGCAATGACGTTGTTTACACCTCGGTTGAATGCCACAG 420
421 ATCTTGAATCTAGAGAACTACAGCAGGTACAGACAGATGGAGAGAGCAGAAG 480
421 ATCTTGAATCTAGAGAACTACAGCAGGTACAGACAGATGGAGAGAGCAGAAG 480
481 CAATGAGAAATCCTTGGGAAAAACACAAACGGAACCTACTAAGAAAGTGAAGAGCAGT 540
481 CAATGAGAAATCCTTGGGAAAAACACAAACGGAACCTACTAAGAAAGTGAAGAGCAGT 540
541 TAGAGTGTGTTGTAACACACTGCATTAAGTGCCTGCTTACTGACCCAGCCCTGC 600
541 TAGAGTGTGTTGTAACACACTGCATTAAGTGCCTGCTTACTGACCCAGCCCTGC 600
601 CAGCACAATCACCTGAGAGCGTCCCATCTCTTATATCTCTCTTCTTCTTCTGCACTG 660
601 CAGCACAATCACCTGAGAGCGTCCCATCTCTTATCTCTCTCTTCTTCTTCTGCACTG 660
661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTGTTTG 720
661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTGTTTG 720
721 ACAGTACTAAGATGTTGAAGTGCCTGCTTGAAGTGCCTGCTTGAAGTGCCTGCTTGA 780
721 ACAGTACTAAGATGTTGAAGTGCCTGCTTGAAGTGCCTGCTTGAAGTGCCTGCTTGA 780
781 CAGTATGAGGGGTACTTGACAAACAGATTAAGAAAGTCCCTGATGCTGCTTAATGCAAG 840
781 CAGTATGAGGGGTACTTGACAAACAGATTAAGAAAGTCCCTGATGCTGCTTAATGCAAG 840
841 ATTGGCTTCTGATGATGTTTCCACCAACATTCAGGAAACAAATCAATTAATAACAA 900
841 ATTGGCTTCTGATGATGTTTCCACCAACATTCAGGAAACAAATCAATTAATAACAA 900
901 AATTGCGAATAATCTTTCTCTAAAGTCAAACTGTTGCAGTTGATTCAGAGGCCAGTG 960
901 AATTGCGAATAATCTTTCTCTAAAGTCAAACTGTTGCAGTTGATTCAGAGGCCAGTG 960
961 TAGCGCTTCTAAATCTGAATAACCCCAAGATTTTCAAGAAATGGAATTAAGCAAACTA 1020
961 TAGCGCTTCTAAATCTGAATAACCCCAAGATTTTCAAGAAATGGAATTAAGCAAACTA 1020
1021 AGAATGACCATGATGAGAAAGAACTGACCATTAAGTCTGTCATATTTATTTGGGATC 1080
1021 AGAATGACCATGATGAGAAAGAACTGACCATTAAGTCTGTCATATTTATTTGGGATC 1080
1081 TGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCGCAGTCTGTTTGAAGATA 1140
1081 TGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTGCGCAGTCTGTTTGAAGATA 1140
1141 ACGAAGACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGTTGCGAATA 1200
1141 ACGAAGACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGTTGCGAATA 1200
1201 TTTTTCATTTGACCAACGGGAGATTTCCATCTGCTGAACTTGCACATCTCTGAGTGA 1260
1201 TTTTTCATTTGACCAACGGGAGATTTCCATCTGCTGAACTTGCACATCTCTGAGTGA 1260
1261 CAATAGTAACACACAGGATGTTTTGAAATTTGAGCCACTTGCCTACCTTTAGTTCCAC 1320

1261 CAATAGTAACACACAGGATGTTTTGAAATTTGAGCCACTTGCCTACCTTTAGTTCCAC 1320
1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAGAAAGTTATTATTACCTAA 1380
1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAGAAAGTTATTATTACCTAA 1380
1381 ATGATGATGTCATGTTTTGGGAAGATGCTGCGCAGATGATTTTACAGTCATCTCCAAAG 1440
1381 ATGATGATGTCATGTTTTGGGAAGATGCTGCGCAGATGATTTTACAGTCATCTCCAAAG 1440
1441 GCCAAGAGTTTATTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCCCT 1500
1441 GCCAAGAGTTTATTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCCCT 1500
1501 GGATTAAGGATGCTATTGTGACAGGCTTGTAAATTAATTCAGCCTCGGATGGATGGTG 1560
1501 GGATTAAGGATGCTATTGTGACAGGCTTGTAAATTAATTCAGCCTCGGATGGATGGTG 1560
1561 GGGATTCCTCTGGAACACAGTGGAGGAGTGCCTATATTGAGGAGGTGGAGGTACTGGGA 1620
1561 GGGATTCCTCTGGAACACAGTGGAGGAGTGCCTATATTGAGGAGGTGGAGGTACTGGGA 1620
1621 GTATTGAGTGGACACCCCTGCGAGTTTGGTGGAGGAATAAACAAGTGTCTCTTACTGTGA 1680
1621 GTATTGAGTGGACACCCCTGCGAGTTTGGTGGAGGAATAAACAAGTGTCTCTTACTGTGA 1680
1681 ATCAGGATGTCGAATTCCTGCTCGCTGATTAAGTTCGTGACCAAGCATGCAATGTCT 1740
1681 ATCAGGATGTCGAATTCCTGCTCGCTGATTAAGTTCGTGACCAAGCATGCAATGTCT 1740
1741 TGCTCTGCTGGTTGATGCTGCGGACTGTGCGGCAAGATCAATTTTCATGAATTTGATAAAG 1800
1741 TGCTCTGCTGGTTGATGCTGCGGACTGTGCGGCAAGATCAATTTTCATGAATTTGATAAAG 1800
1801 TGATCTTCTCCCAACCCAGACTCACTATATTATTCAAAAGTGAATGCTCTGCTTATT 1860
1801 TGATCTTCTCCCAACCCAGACTCACTATATTATTCAAAAGTGAATGCTCTGCTTATT 1860
1861 TCAGCTTTGAGAGTGTAGCCAAAGAGGAGTTCGAAGTGCCTATAGTGAACAATCCAAATA 1920
1861 TCAGCTTTGAGAGTGTAGCCAAAGAGGAGTTCGAAGTGCCTATAGTGAACAATCCAAATA 1920
1921 TTGACATGCTTCTATTGCGCAACAGTGGAAACCATCCACCTCATTAATGACAGTGGAA 1980
1921 TTGACATGCTTCTATTGCGCAACAGTGGAAACCATCCACCTCATTAATGACAGTGGAA 1980
1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATAACAAACGATGAAGATTCA 2040
1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATAACAAACGATGAAGATTCA 2040
2041 AAATGACAGATAACAGTGGAGGTGGAACAAGGAGGGAACAAACTGAAATTTCTACGGCCC 2100
2041 AAATGACAGATAACAGTGGAGGTGGAACAAGGAGGGAACAAACTGAAATTTCTACGGCCC 2100
2101 AGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGCGGAATCCTTT 2160
2101 AGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGCGGAATCCTTT 2160
2161 TTGAGGATATTTCCCAAGAAAAACGGTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220
2161 TTGAGGATATTTCCCAAGAAAAACGGTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220
2221 CAAGGAGAGCCAGGAGGTGAAAATTTCCCTCGTGAATAATTTCTCTCTTTCCAAAAG 2280
2221 CAAGGAGAGCCAGGAGGTGAAAATTTCCCTCGTGAATAATTTCTCTCTTTCCAAAAG 2280
2281 ACGCCAGTTGAGTCTCAATACCTCGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340
2281 ACGCCAGTTGAGTCTCAATACCTCGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340
2341 AAGGATACAAATTTGTCAGAGTCAAGTCTGCTGAGATCAATTTCTGATGAATCAACAGATG 2400
2341 AAGGATACAAATTTGTCAGAGTCAAGTCTGCTGAGATCAATTTCTGATGAATCAACAGATG 2400

QY	781	CAGTATGAGGGGGTACTTGAACAACAGATGAAGAGTCCCTGGATTAGTCTAATGAAG	840
DB	781	CAGTATGAGGGGGTACTTGAACAACAGATGAAGAGTCCCTGGATTAGTCTAATGAAG	840
QY	841	ATTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAARACAA	900
DB	841	ATTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAARACAA	900
QY	901	AATTGCCAGAAAATCTTTCCCTTAAGAGTCAAACTGTTGCAAGTCTAATTCAGAGGCCAGTG	960
DB	901	AATTGCCAGAAAATCTTTCCCTTAAGAGTCAAACTGTTGCAAGTCTAATTCAGAGGCCAGTG	960
QY	961	TAGCGCTTCTAAAATGMAATACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAGA	1020
DB	961	TAGCGCTTCTAAAATGMAATACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAGA	1020
QY	1021	AGAACATGACCATTTGATGGAAGAACTGACCAATAGTCTGCAATATTATTATGGGATC	1080
DB	1021	AGAACATGACCATTTGATGGAAGAACTGACCAATAGTCTGCAATATTATTATGGGATC	1080
QY	1081	TGAGCGCCATCAGCGAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTTGAAGATA	1140
DB	1081	TGAGCGCCATCAGCGAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTTGAAGATA	1140
QY	1141	ACGAGAACTGAGGTACTCATTTGGATCTATCGAGAGCATGACCATGGSTTCGGATA	1200
DB	1141	ACGAGAACTGAGGTACTCATTTGGATCTATCGAGAGCATGACCATGGSTTCGGATA	1200
QY	1201	TTTTCAATGTCAACCAAGGGAGATTCATCTCTGGCTGAACCTTGACAACTCTCGAGTGA	1260
DB	1201	TTTTCAATGTCAACCAAGGGAGATTCATCTCTGGCTGAACCTTGACAACTCTCGAGTGA	1260
QY	1261	CAATAGTAACACACAGAGATGTTTTGCAATTTGAGCCACTTGCCTACCTTTAGTTTAC	1320
DB	1261	CAATAGTAACACACAGAGATGTTTTGCAATTTGAGCCACTTGCCTACCTTTAGTTTAC	1320
QY	1321	CTGCTATTGAAAGTCACATTCATCGCATGAGGGCTGTCGAGAGTATTATTACCTAA	1380
DB	1321	CTGCTATTGAAAGTCACATTCATCGCATGAGGGCTGTCGAGAGTATTATTACCTAA	1380
QY	1381	ATGATGATGTCATGTTTGGGAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG	1440
DB	1381	ATGATGATGTCATGTTTGGGAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG	1440
QY	1441	CCGAGAGGTTTATTGACATGGCTGTCGCAAACTGTCGAGGGCTGCCAGGTTTCT	1500
DB	1441	CCGAGAGGTTTATTGACATGGCTGTCGCAAACTGTCGAGGGCTGCCAGGTTTCT	1500
QY	1501	GGATTAAAGTGGCTTATTGACAGGCTTGTAAATTAATTCAGCTGCGANTGGGATGGTG	1560
DB	1501	GGATTAAAGTGGCTTATTGACAGGCTTGTAAATTAATTCAGCTGCGANTGGGATGGTG	1560
QY	1561	GGGATGCTCGGAACAGTGGAGGAGTCCCTATATTGCAAGAGGTGGAGTACTGGGA	1620
DB	1561	GGGATGCTCGGAACAGTGGAGGAGTCCCTATATTGCAAGAGGTGGAGTACTGGGA	1620
QY	1621	GTAATGGAGTTGGAACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA	1680
DB	1621	GTAATGGAGTTGGAACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA	1680
QY	1681	ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTCTGTGACCAAGCATGCAATGCT	1740
DB	1681	ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTCTGTGACCAAGCATGCAATGCT	1740
QY	1741	TGTCCTGTGGGTTTGTATGCTGGGACATGTCGGAAGATCATTTTCATGAATTTGATAAAG	1800
DB	1741	TGTCCTGTGGGTTTGTATGCTGGGACATGTCGGAAGATCATTTTCATGAATTTGATAAAG	1800
QY	1801	TGATCCTTCTCCAAACAGACTCATATATTATTCMAAAGGTGAATGCTGCTTATT	1860
DB	1801	TGATCCTTCTCCAAACAGACTCATATATTATTCMAAAGGTGAATGCTGCTTATT	1860

QY	1861	TCAGCTTTGCAGAGTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGCACAAATCCAATAA	1920
DB	1861	TCAGCTTTGCAGAGTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGCACAAATCCAATAA	1920
QY	1921	TTTCGACATGCTTCTATTGCCAACAGTGGAAAACCATCCACTCATATATGCACAGTGAA	1980
DB	1921	TTTCGACATGCTTCTATTGCCAACAGTGGAAAACCATCCACTCATATATGCACAGTGAA	1980
QY	1981	TGAATGCCACCAACAATATATTTTAAATCTCAGGTTTCAAAAATACAAAAGATGAAGTCA	2040
DB	1981	TGAATGCCACCAACAATATATTTTAAATCTCAGGTTTCAAAAATACAAAAGATGAAGTCA	2040
QY	2041	AAATCGAGATAACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACGGGCC	2100
DB	2041	AAATCGAGATAACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACGGGCC	2100
QY	2101	AGAAGGGTTACGAAAATTTAGTTAGTCCCAATAACAATTTCTCCAGAGCGGAAATTCCTTT	2160
DB	2101	AGAAGGGTTACGAAAATTTAGTTAGTCCCAATAACAATTTCTCCAGAGCGGAAATTCCTTT	2160
QY	2161	TTGAGGATATTTCCCAAGAAAACGCTTCCCGAGTTTAAAGAGACATGATTTAACTCAA	2220
DB	2161	TTGAGGATATTTCCCAAGAAAACGCTTCCCGAGTTTAAAGAGACATGATTTAACTCAA	2220
QY	2221	CAAGGAGAGCCAGGAGAGGAGTGAATAATTCCTGGTAAATTTCTACTCTCTCCAAAG	2280
DB	2221	CAAGGAGAGCCAGGAGAGGAGTGAATAATTCCTGGTAAATTTCTACTCTCTCCAAAG	2280
QY	2281	AGCCGAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCACTTTGA	2340
DB	2281	AGCCGAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCACTTTGA	2340
QY	2341	AAGGATACAAATTTGCTCAAGTCAGCTTCTGATGATCAATTTCTGATGAATTCACAGCATG	2400
DB	2341	AAGGATACAAATTTGCTCAAGTCAGCTTCTGATGATCAATTTCTGATGAATTCACAGCATG	2400
QY	2401	CTAAAATAAAATCAAGCTATAAACAAGTGAATAAACAATTAACAGTGGTGGCTCCAC	2460
DB	2401	CTAAAATAAAATCAAGCTATAAACAAGTGAATAAACAATTAACAGTGGTGGCTCCAC	2460
QY	2461	AGGAAACAGGTTTCAATAAAGCATCTTGCACAAACAGCTTAGGAGTGTCTCAAGATTGC	2520
DB	2461	AGGAAACAGGTTTCAATAAAGCATCTTGCACAAACAGCTTAGGAGTGTCTCAAGATTGC	2520
QY	2521	AGAGGTTGATCTTTCTGAGTGAATAAGTGAATGGTCAAGCAGGCTCAGAAATC	2580
DB	2521	AGAGGTTGATCTTTCTGAGTGAATAAGTGAATGGTCAAGCAGGCTCAGAAATC	2580
QY	2581	CACCCCTGGACTTGGAGACCAAGATTTAGAGTGGAACTCACAACCAAAACCA	2640
DB	2581	CACCCCTGGACTTGGAGACCAAGATTTAGAGTGGAACTCACAACCAAAACCA	2640
QY	2641	TAGGCGGAAATGTGACAAAAGAAAAGCCCCATCTCTGATTGTTCCACTGGAAAGCCAGA	2700
DB	2641	TAGGCGGAAATGTGACAAAAGAAAAGCCCCATCTCTGATTGTTCCACTGGAAAGCCAGA	2700
QY	2701	TGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAAAAGACAGTAGATGGAGGAAATG	2760
DB	2701	TGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAAAAGACAGTAGATGGAGGAAATG	2760
QY	2761	CTGAAAATCACATAGGCGTTTACTGAAAGTGTACTTTGGAAGAAAAGTGCAGCATTAACAG	2820
DB	2761	CTGAAAATCACATAGGCGTTTACTGAAAGTGTACTTTGGAAGAAAAGTGCAGCATTAACAG	2820
QY	2821	ATAGTTACTTGGGCTTTTGGCATGGGAAAAGAAAAGTATTTCCAGATCTTCTCGACG	2880
DB	2821	ATAGTTACTTGGGCTTTTGGCATGGGAAAAGAAAAGTATTTCCAGATCTTCTCGACG	2880
QY	2881	AAGAAGGTCATTGAAGACAAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC	2940
DB	2881	AAGAAGGTCATTGAAGACAAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC	2940
QY	2941	AACTAAAAGATACATTTTCAGATTCCTCTCAGATATCTAAATAAAATTTCTAAATAGCAAGT	3000

Db 2941 AACTAAAGATACATTTGACAGATCCCTCAGATATGTAATAAATTTCTAATAAGCAAGT 3000
Qy 3001 TTGGATTACATCGGGAAGTCCCTGCTCAGTCCCTCAGATGATGACCGGATTGTTA 3060
Db 3001 TTGGATTACATCGGGAAGTCCCTGCTCAGTCCCTCAGATGATGACCGGATTGTTA 3060
Qy 3061 TCGAAGAACTCAAGATATGTTCCCTGGAAGAAATTTGACAGACGTCAATTTACAAAGTGC 3120
Db 3061 TCGAAGAACTCAAGATATGTTCCCTGGAAGAAATTTGACAGACGTCAATTTACAAAGTGC 3120
Qy 3121 GCCATTCTGAGGATATGAGATTTGCCCTTCTCTTATTTTATATCTCATGAGTGCAGTGC 3180
Db 3121 GCCATTCTGAGGATATGAGATTTGCCCTTCTCTTATTTTATATCTCATGAGTGCAGTGC 3180
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAGTGTGATACAGATCAATCTGTTGCTTGT 3240
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAGTGTGATACAGATCAATCTGTTGCTTGT 3240
Qy 3241 CTGACAGAAATCCGAACACTGGCTACCGAAATTCACGAATCGCCGTTAAGTTTGCAGG 3300
Db 3241 CTGACAGAAATCCGAACACTGGCTACCGAAATTCACGAATCGCCGTTAAGTTTGCAGG 3300
Qy 3301 ATTTGAGAGTCTGGACACATGCTAATAAATTTGCTCAAAAATGCTTCTCTGCTGATATCA 3360
Db 3301 ATTTGAGAGTCTGGACACATGCTAATAAATTTGCTCAAAAATGCTTCTCTGCTGATATCA 3360
Qy 3361 CGCAGCTAAATTAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420
Db 3361 CGCAGCTAAATTAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420
Qy 3421 TCACATAAGTCTAGTAACAACTGTAAACAGTAATCTGACAAATCCCAAGAGCATATA 3480
Db 3421 TCACATAAGTCTAGTAACAACTGTAAACAGTAATCTGACAAATCCCAAGAGCATATA 3480
Qy 3481 AGGCAAAAACAAATATAGTTTGAATCATGGSAGAGAGAAATCGCTTTTAAATGA 3540
Db 3481 AGGCAAAAACAAATATAGTTTGAATCATGGSAGAGAGAAATCGCTTTTAAATGA 3540
Qy 3541 TTGCTACCAACGTTTCTCATGTGTTGGCAGTTGGATGACATAAGAAAAACCCTAGGA 3600
Db 3541 TTGCTACCAACGTTTCTCATGTGTTGGCAGTTGGATGACATAAGAAAAACCCTAGGA 3600
Qy 3601 AGTTGTTTCCCTGGAATGACAACTTGCACCAATCATAAAGATGCTCAGACAGTGAAG 3660
Db 3601 AGTTGTTTCCCTGGAATGACAACTTGCACCAATCATAAAGATGCTCAGACAGTGAAG 3660
Qy 3661 CTGTTCTCAGGACTTCTATGATCCATGTTCCCATACCTTCCCAATTTGAATGCCAA 3720
Db 3661 CTGTTCTCAGGACTTCTATGATCCATGTTCCCATACCTTCCCAATTTGAATGCCAA 3720
Qy 3721 GAGAGTATCGAAACCGTTTCTCATATGATGATGAGTGCAGGATGAGGCTTTATCGAG 3780
Db 3721 GAGAGTATCGAAACCGTTTCTCATATGATGATGAGTGCAGGATGAGGCTTTATCGAG 3780
Qy 3781 ACAAAATGAAATTTTGGACCATTTGCTAGTACGACATGATGATGTTTACTATATCT 3840
Db 3781 ACAAAATGAAATTTTGGACCATTTGCTAGTACGACATGATGATGTTTACTATATCT 3840
Qy 3841 CATTTTTCAGCAGTAAATGCACTTAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900
Db 3841 CATTTTTCAGCAGTAAATGCACTTAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900
Qy 3901 ACAAGAGCTAGTCCCAATCGATCAGATATAGAGATCTTCATTTGAAACCAATCTA 3960
Db 3901 ACAAGAGCTAGTCCCAATCGATCAGATATAGAGATCTTCATTTGAAACCAATCTA 3960
Qy 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTGAT 4020
Db 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTGAT 4020
Qy 4021 GCTTAGCAGTTTGGCCGGAAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080

Db 4021 GCTTAGCAGTTTGGCCGGAAGAGAAATATATCCAGTACCATGCTGTTTGTGGCATGAA 4080
Qy 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140
Db 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140
Qy 4141 TCTGAACTCATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTCATCTTTTATAAAA 4200
Db 4141 TCTGAACTCATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTCATCTTTTATAAAA 4200
Qy 4201 GGCTCACTGACAGAGACAGCTGTTTAAATTTCCACAGCAATCATTTGCAGACTTAACTTAT 4260
Db 4201 GGCTCACTGACAGAGACAGCTGTTTAAATTTCCACAGCAATCATTTGCAGACTTAACTTAT 4260
Qy 4261 TAGGAGAACCTATGCCAGCTGGGAGTATGCTTAAGAGGCTCCAGTCTTTGCCATTCMA 4320
Db 4261 TAGGAGAACCTATGCCAGCTGGGAGTATGCTTAAGAGGCTCCAGTCTTTGCCATTCMA 4320
Qy 4321 AGCCTTTTCTCTAAAGTTTTCACACTTTTTCATTTCCCATTTTAAAGTAGTACTA 4380
Db 4321 AGCCTTTTCTCTAAAGTTTTCACACTTTTTCATTTCCCATTTTAAAGTAGTACTA 4380
Qy 4381 AGTTAACTAGTTTATTTCTGCTTCTGAGTATAACGAATTTGGGATGCTTAAACCTATTTTA 4440
Db 4381 AGTTAACTAGTTTATTTCTGCTTCTGAGTATAACGAATTTGGGATGCTTAAACCTATTTTA 4440
Qy 4441 TAGATGTTTATTTAAATAATGACAGCAATATCACCTCTTATTGACAATACCTAAATATGAG 4500
Db 4441 TAGATGTTTATTTAAATAATGACAGCAATATCACCTCTTATTGACAATACCTAAATATGAG 4500
Qy 4501 TTTTATTAATATTTAAGACTGPAATGCTTTTAAACCACTAACTACTGAAGAGCTCAATG 4560
Db 4501 TTTTATTAATATTTAAGACTGPAATGCTTTTAAACCACTAACTACTGAAGAGCTCAATG 4560
Qy 4561 ATTGACATCTGAAATGCTTTGTAATTTGACTTCAGCCCTTAAAGATGCTATGATTTCA 4620
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTGACTTCAGCCCTTAAAGATGCTATGATTTCA 4620
Qy 4621 CGTGAGGCTTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT 4680
Db 4621 CGTGAGGCTTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT 4680
Qy 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGATGAGTACTGCTGATGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCATTTTAAAGATGAGTACTGCTGATGCA 4740
Qy 4741 CTGGCACAATACAGTGGTGTCTTTTAACTCACTGCTATATTCCAGTAGTACTCTCT 4800
Db 4741 CTGGCACAATACAGTGGTGTCTTTTAACTCACTGCTATATTCCAGTAGTACTCTCT 4800
Qy 4801 CTCAGTTGGTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860
Db 4801 CTCAGTTGGTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860
Qy 4861 AAATTTATTCAGGCCACCTGCTCTTTGTCATACATCTTCTGCTGTTTGTAGTTTGT 4920
Db 4861 AAATTTATTCAGGCCACCTGCTCTTTGTCATACATCTTCTGCTGTTTGTAGTTTGT 4920
Qy 4921 TTTTTCCTCAAAACCCCTTAAATAATGTAATAACCAATGTTAGCTTGCAGCTGTACAAA 4980
Db 4921 TTTTTCCTCAAAACCCCTTAAATAATGTAATAACCAATGTTAGCTTGCAGCTGTACAAA 4980
Qy 4981 ACTGCCACCCAGCAGATGACCCCTCAGGCCATCATTTGCCAATCATCTGAGAATTTT 5040
Db 4981 ACTGCCACCCAGCAGATGACCCCTCAGGCCATCATTTGCCAATCATCTGAGAATTTT 5040
Qy 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100
Qy 5101 TGCAGTGGCGCAATCTCAGCTCAGTCACTGCAACCTCCGCTCCCGGTTTCAAGCAGTTCTGTC 5160
Db 5101 TGCAGTGGCGCAATCTCAGCTCAGTCACTGCAACCTCCGCTCCCGGTTTCAAGCAGTTCTGTC 5160

QY 1321 CTGCTATTGAAGTCACTTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTACCTAA 1380
Db 1321 CTGCTATTGAAGTCACTTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTACCTAA 1380
QY 1381 ATGATGATGATCATCTTTTGGGAAGATCTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
Db 1381 ATGATGATGATCATCTTTTGGGAAGATCTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440
QY 1441 GCCAAGAGGTTTATTTGACATGGGCTGTGCCAACTGTGCCAGGGGTGCCAGGTTCCCT 1500
Db 1441 GCCAAGAGGTTTATTTGACATGGGCTGTGCCAACTGTGCCAGGGGTGCCAGGTTCCCT 1500
QY 1501 GGATTAAGGATGCTATTGTGACAAAGCTTTGTAATAATTACAGCTCGGATGGATGGTG 1560
Db 1501 GGATTAAGGATGCTATTGTGACAAAGCTTTGTAATAATTACAGCTCGGATGGATGGTG 1560
QY 1561 GGGATTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGGTACTGGGA 1620
Db 1561 GGGATTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGGTACTGGGA 1620
QY 1621 GTATTGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTGA 1680
Db 1621 GTATTGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTGA 1680
QY 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740
Db 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740
QY 1741 TGTCTGTGGTTTGTAGTCTGGGACCTGTGGGCAAGATCAATTTTCATGAATTTGATAAAG 1800
Db 1741 TGTCTGTGGTTTGTAGTCTGGGACCTGTGGGCAAGATCAATTTTCATGAATTTGATAAAG 1800
QY 1801 TGATCTCTTCCCAAACAGACTCACTATATTTATTCAAAAGTGAATGCTGCCCTATT 1860
Db 1801 TGATCTCTTCCCAAACAGACTCACTATATTTATTCAAAAGTGAATGCTGCCCTATT 1860
QY 1861 TCAGCTTTGAGAAAGTGAAGAGGAGTGAAGTGCCTATAGTGACAAATCCCAATAA 1920
Db 1861 TCAGCTTTGAGAAAGTGAAGAGGAGTGAAGTGCCTATAGTGACAAATCCCAATAA 1920
QY 1921 TTGACATGCTTTTATTTGCCAAACAAGTGGAAAACCACTCACTCATATGCAAGTGGAA 1980
Db 1921 TTGACATGCTTTTATTTGCCAAACAAGTGGAAAACCACTCACTCATATGCAAGTGGAA 1980
QY 1981 TGAATGCCACACATACTTTTAATCTCACGTTTCAAAATACAAACGATGAAGATTCA 2040
Db 1981 TGAATGCCACACATACTTTTAATCTCACGTTTCAAAATACAAACGATGAAGATTCA 2040
QY 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACTCAATTTCTACGGCCC 2100
Db 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACTCAATTTCTACGGCCC 2100
QY 2101 AGRAGGTTACGAAATTTTAGTTAGTCCCATACACTTCTTCCAGAGGGGGAATCCTTT 2160
Db 2101 AGRAGGTTACGAAATTTTAGTTAGTCCCATACACTTCTTCCAGAGGGGGAATCCTTT 2160
QY 2161 TTGAGGATATTTCCAAAGAAAACCGCTTCCGGAAGTTTAAGACATGATTTAACTCAA 2220
Db 2161 TTGAGGATATTTCCAAAGAAAACCGCTTCCGGAAGTTTAAGACATGATTTAACTCAA 2220
QY 2221 CAAGGAGAGCCAGGAGAGGTGAATAATTCCTCGTGAATAATTTCACTCTCCAAAAG 2280
Db 2221 CAAGGAGAGCCAGGAGAGGTGAATAATTCCTCGTGAATAATTTCACTCTCTCCAAAAG 2280
QY 2281 ACGCCAGTTGAGTCTCAATACCTTGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340
Db 2281 ACGCCAGTTGAGTCTCAATACCTTGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340
QY 2341 AAGGATACAAATTTGTCAGTCAAGCTTGTGTGATCATTTCTGATGAATCAAGCATG 2400
Db 2341 AAGGATACAAATTTGTCAGTCAAGCTTGTGTGATCATTTCTGATGAATCAAGCATG 2400
QY 2401 CTAATAATAAATCAAGCTTATAATACATGAAACAAATGACAGTTTGGTGGCTCCAC 2460

Db 2401 CTAATAATAAATCAAGCTTATAATACATGAAACAAATGACAGTTTGGTGGCTCCAC 2460
QY 2461 AGRAGGAGAGGTTTCAATTAAGCACTTTCGCAACAGCTTGGAGTGTCTGAAGATTGC 2520
Db 2461 AGRAGGAGAGGTTTCAATTAAGCACTTTCGCAACAGCTTGGAGTGTCTGAAGATTGC 2520
QY 2521 AGAGGTTGACTTTTCTCTGAGTGTAAAGTGAATGGTCAAGCAGGTCAGAAATC 2580
Db 2521 AGAGGTTGACTTTTCTCTGAGTGTAAAGTGAATGGTCAAGCAGGTCAGAAATC 2580
QY 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGAAGCACTCACACCCAAAACCA 2640
Db 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGAAGCACTCACACCCAAAACCA 2640
QY 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700
Db 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700
QY 2701 TGACAAAAGAAAAGAAAATCAAGGGAAGAAAAGAGAACAGTGAATGGAGGAAAATG 2760
Db 2701 TGACAAAAGAAAAGAAAATCAAGGGAAGAAAAGAGAACAGTGAATGGAGGAAAATG 2760
QY 2761 CTGAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACAG 2820
Db 2761 CTGAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACAG 2820
QY 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAGATTTTCCAAAGATCTTCTCGAGC 2880
Db 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAGATTTTCCAAAGATCTTCTCGAGC 2880
QY 2881 AGRAGAGTCAATGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATCTCGGAGGC 2940
Db 2881 AGRAGAGTCAATGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATCTCGGAGGC 2940
QY 2941 AACTAAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
Db 2941 AACTAAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000
QY 3001 TTGGATTCATATCGCGGAAAAGTCCCTGCTCACTGCTCAATGATGACCGGATTTGTTA 3060
Db 3001 TTGGATTCATATCGCGGAAAAGTCCCTGCTCACTGCTCAATGATGACCGGATTTGTTA 3060
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGTCAATTTCAAAAGTGC 3120
Db 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGTCAATTTCAAAAGTGC 3120
QY 3121 GCCATCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180
Db 3121 GCCATCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180
QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240
QY 3241 CTGACAGAGAAATCCGAACTGCTACAGAAATTCAGAACTGCGCTAAAGTTTGCAGG 3300
Db 3241 CTGACAGAGAAATCCGAACTGCTACAGAAATTCAGAACTGCGCTAAAGTTTGCAGG 3300
QY 3301 ATTTGACAGGCTGGAACACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
Db 3301 ATTTGACAGGCTGGAACACATGCTAATAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360
QY 3361 GCGAGTAAATAATTTCCACCACTCAGGAAATCTTACTATGATCCCAACCTGCCACCGG 3420
Db 3361 GCGAGTAAATAATTTCCACCACTCAGGAAATCTTACTATGATCCCAACCTGCCACCGG 3420
QY 3421 TCATTAAGTCTAGTAAACAACTGTAAAACAGTAACTGACAAAATCCAAAAGCATATA 3480
Db 3421 TCATTAAGTCTAGTAAACAACTGTAAAACAGTAACTGACAAAATCCAAAAGCATATA 3480
QY 3481 AGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAAGATGCTTTTAAATGA 3540


```
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-306-686-4

Query Match      100.0%; Score 5597; DB 15; Length 5597;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCCGAGCGGCGTCCGTGCGCGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60
DB 1 CGGAGCCGAGCGGCGTCCGTGCGCGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60
QY 61 TGC CGCGCGGCGCCCGGAGCTGCAATGAGCGCGCTCGTGAAGCGCGCGCGCGG 120
DB 61 TGC CGCGCGGCGCCCGGAGCTGCAATGAGCGCGCTCGTGAAGCGCGCGCGG 120
QY 121 GCGCGCTCAGGCTCCTCGCGGCGTGGCGTGGCGTGAAGGCTGATGCTCTCAAGCTCC 180
DB 121 GCGCGCTCAGGCTCCTCGCGGCGTGGCGTGGCGTGAAGGCTGATGCTCTCAAGCTCC 180
QY 181 TGCAGAGACAAACCTATACCTGCTGCTGCCACAGGATGCGCTACGTTCTTTGG 240
DB 181 TGCAGAGACAAACCTATACCTGCTGCTGCCACAGGATGCGCTACGTTCTTTGG 240
QY 241 GCGTGTGTCACCATGCTCCGCTTCCAGTTCGAGAGGCTGCTTGGATGAGCC 300
DB 241 GCGTGTGTCACCATGCTCCGCTTCCAGTTCGAGAGGCTGCTTGGATGAGCC 300
QY 301 GAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAGTCCCTTC 360
DB 301 GAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAGTCCCTTC 360
QY 361 AGAATCGGCTTGTCTGCCATGCCGATTGACGTTGTTACCTGCGGTGAATGGCACAG 420
DB 361 AGAATCGGCTTGTCTGCCATGCCGATTGACGTTGTTACCTGCGGTGAATGGCACAG 420
QY 421 ATCTTGAACTACTGAAGGAACCTACAGCAGGTCAGAGAACAGATGGAGGAGCAGAAAG 480
DB 421 ATCTTGAACTACTGAAGGAACCTACAGCAGGTCAGAGAACAGATGGAGGAGCAGAAAG 480
QY 481 CAATGAGAGAAATCCTTTGGGAAACACACCGAAACCTACTAAGAGAGTGAAGCAGT 540
DB 481 CAATGAGAGAAATCCTTTGGGAAACACACCGAAACCTACTAAGAGAGTGAAGCAGT 540
QY 541 TAGAGTGTGCTAACACACTGCATTAAGTGGCCATGCTGTACTGGACCCAGCCCTGC 600
DB 541 TAGAGTGTGCTAACACACTGCATTAAGTGGCCATGCTGTACTGGACCCAGCCCTGC 600
QY 601 CAGCCAAACATCACCTGAGGAGCGTCCATCTCTTTATCTCTTTTCAATTCGCCAGTG 660
DB 601 CAGCCAAACATCACCTGAGGAGCGTCCATCTCTTTATCTCTTTTCAATTCGCCAGTG 660
QY 661 ACATTTCAATGTTGAAAACCAAAAACCTTCTACCAATGCTCAGTTGTTGTTT 720
DB 661 ACATTTCAATGTTGAAAACCAAAAACCTTCTACCAATGCTCAGTTGTTGTTT 720
QY 721 ACAGTACTAAGGATGTTGAAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGACAGA 780
DB 721 ACAGTACTAAGGATGTTGAAGATGCCACCTCTGACCTGCTTAAAGGAAATAGCAGACAGA 780
QY 781 CAGTATGAGGGGGTACTTTGCAACAGATATAAGAGTCCCTGGATTAGTGTCTTAATGCAAG 840
DB 781 CAGTATGAGGGGGTACTTTGCAACAGATATAAGAGTCCCTGGATTAGTGTCTTAATGCAAG 840
QY 841 ATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAACAA 900
DB 841 ATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAACAA 900
QY 901 AATTGCCAGAAAATCTTCTCTTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCGCAGTG 960
DB 901 AATTGCCAGAAAATCTTCTCTTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCGCAGTG 960
QY 961 TAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTTCAAGAAATGGAATAAGCAAACTAAGA 1020
DB 961 TAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTTCAAGAAATGGAATAAGCAAACTAAGA 1020
QY 1021 AGAATGACCATTCATGAGAAAGAACTGACATAAGTCCCTGCATTTATTTATGGGATC 1080
DB 1021 AGAATGACCATTCATGAGAAAGAACTGACATAAGTCCCTGCATTTATTTATGGGATC 1080
QY 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA 1140
DB 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA 1140
QY 1141 ACGAAGAACTGAGGTACTTATTCGATCTATTCGAGAGCATGCCACATGCGGTTCGGAATA 1200
DB 1141 ACGAAGAACTGAGGTACTTATTCGATCTATTCGAGAGCATGCCACATGCGGTTCGGAATA 1200
QY 1201 TTTTCATTTGTCACCAACCGGAGATTCATCTCTGCTGAACTCTTGACAACTCTCCAGTGA 1260
DB 1201 TTTTCATTTGTCACCAACCGGAGATTCATCTCTGCTGAACTCTTGACAACTCTCCAGTGA 1260
QY 1261 CAATAGTAACACACACGAGGATGTTTTCGAAATTTGAGCCACTTGCTTACCTTAC 1320
DB 1261 CAATAGTAACACACACGAGGATGTTTTCGAAATTTGAGCCACTTGCTTACCTTAC 1320
QY 1321 CTGCTATTGAAAGTCAATTCATTCGATCGAAGGCTGTCCAGAGGTTTATTTACCTAA 1380
DB 1321 CTGCTATTGAAAGTCAATTCATTCGATCGAAGGCTGTCCAGAGGTTTATTTACCTAA 1380
QY 1381 ATGATGATGTCATGTTTGGGAGGATGTCGCGCCAGATGATTTTACAGTCACTCCAAAG 1440
DB 1381 ATGATGATGTCATGTTTGGGAGGATGTCGCGCCAGATGATTTTACAGTCACTCCAAAG 1440
QY 1441 GCGAGAGGTTTATTTGACATGGCTGTGCCAACTGTGCCGAGGCTGCCAGGTTCCCT 1500
DB 1441 GCGAGAGGTTTATTTGACATGGCTGTGCCAACTGTGCCGAGGCTGCCAGGTTCCCT 1500
QY 1501 GGATTAAGGATGGCTATTGTCACAGGCTTGTAAATAATTACGCTGCCATTGGGATGCTG 1560
DB 1501 GGATTAAGGATGGCTATTGTCACAGGCTTGTAAATAATTACGCTGCCATTGGGATGCTG 1560
QY 1561 GGGATTGCTCTGGAACAGTGGAGGAGTCCCTATATTTCGAGGAGGTTGAGGTACTGGGA 1620
DB 1561 GGGATTGCTCTGGAACAGTGGAGGAGTCCCTATATTTCGAGGAGGTTGAGGTACTGGGA 1620
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAAATAAACAGTGTCTTCTACTGTA 1680
DB 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAAATAAACAGTGTCTTCTACTGTA 1680
QY 1681 ATCAGGAGTGGCGAATTCCTGGCTGCTGATTAAGTCTGTGACCAAGATGCAATGTCT 1740
DB 1681 ATCAGGAGTGGCGAATTCCTGGCTGCTGATTAAGTCTGTGACCAAGATGCAATGTCT 1740
QY 1741 TGTCCTGTTGGTTGATCTGGCGACTGTGGCAGATCAATTTTCATGAATGTTATAAG 1800
DB 1741 TGTCCTGTTGGTTGATCTGGCGACTGTGGCAGATCAATTTTCATGAATGTTATAAG 1800
QY 1801 TGATCTCTTCTCCCAACCCAGACTCACTATATTATTTCCAAAAGGTGAATGCTGCTTATT 1860
DB 1801 TGATCTCTTCTCCCAACCCAGACTCACTATATTATTTCCAAAAGGTGAATGCTGCTTATT 1860
```

1861 TCAGCTTTGAGAGTAGCCCAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCAAATA 1920 QY
1861 TCAGCTTTGAGAGTAGCCCAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCAAATA 1920 Db
1921 TTGACATGCTTCTATTGCGCAACAAAGTGGAAACCAATCCACCTCAPAATGSCAGTGGAA 1980 QY
1921 TTGACATGCTTCTATTGCGCAACAAAGTGGAAACCAATCCACCTCAPAATGSCAGTGGAA 1980 Db
1981 TGAATGCCACCAATACATTTTAACTCTACGTTTCAAAATCAACAGATGAGAGTTCA 2040 QY
1981 TGAATGCCACCAATACATTTTAACTCTACGTTTCAAAATCAACAGATGAGAGTTCA 2040 Db
2041 AAATGACAGATAACAGTGGAGTGGACACAAAGGAGGAGCAAAACTGAAATTTCTACGGCCC 2100 QY
2041 AAATGACAGATAACAGTGGAGTGGACACAAAGGAGGAGCAAAACTGAAATTTCTACGGCCC 2100 Db
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATCCCTTT 2160 QY
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATCCCTTT 2160 Db
2161 TTGAGGATATTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220 QY
2161 TTGAGGATATTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220 Db
2221 CAAGGAGAGCCAGGAGAGGTGAAAATTTCCCTGTTAAATATTTCACTCTCTTCCAAAAG 2280 QY
2221 CAAGGAGAGCCAGGAGAGGTGAAAATTTCCCTGTTAAATATTTCACTCTCTTCCAAAAG 2280 Db
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCACTGGAACATGAGACATCATCTTGA 2340 QY
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCACTGGAACATGAGACATCATCTTGA 2340 Db
2341 AAGGATACAAATTTGTCACAGTACGCTTGTGAGATCATTTCTGATGAACTCACAGCATG 2400 QY
2341 AAGGATACAAATTTGTCACAGTACGCTTGTGAGATCATTTCTGATGAACTCACAGCATG 2400 Db
2401 CTAAATTAATAATCAAGCTATTAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460 QY
2401 CTAAATTAATAATCAAGCTATTAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460 Db
2461 AGGAAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAAAGTTGC 2520 QY
2461 AGGAAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAAAGTTGC 2520 Db
2521 AGAGTTGACTTTTCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2580 QY
2521 AGAGTTGACTTTTCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2580 Db
2581 CACCCCTGAGCTTGGAGACACAGCAAGATTAGAGTGGAAATCTCACACCCCAAAACCA 2640 QY
2581 CACCCCTGAGCTTGGAGACACAGCAAGATTAGAGTGGAAATCTCACACCCCAAAACCA 2640 Db
2641 TAGCCGAAATGTGACAAAAGAAAAGCCCAATCTCTGATGTTGTTCCACTGGAAGCCAGA 2700 QY
2641 TAGCCGAAATGTGACAAAAGAAAAGCCCAATCTCTGATGTTGTTCCACTGGAAGCCAGA 2700 Db
2701 TGACAAAGAAAAGAAAATCACAGGGAAGAAAAGAACAGACAGTGAATGGAGGAAAATG 2760 QY
2701 TGACAAAGAAAAGAAAATCACAGGGAAGAAAAGAACAGACAGTGAATGGAGGAAAATG 2760 Db
2761 CTGAAAATCATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGATTTACACAG 2820 QY
2761 CTGAAAATCATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGATTTACACAG 2820 Db
2821 ATAGTTACTTGGGCTTTTCCATGGAGAAAAGAAAGTATTTCCAGATCTTCTCGAGC 2880 QY
2821 ATAGTTACTTGGGCTTTTCCATGGAGAAAAGAAAGTATTTCCAGATCTTCTCGAGC 2880 Db
2881 AAGAGAGTCAATGAGACACAAATGGCATACTTCACTGATAGCAAAAATCTGGGAGGC 2940 QY
2881 AAGAGAGTCAATGAGACACAAATGGCATACTTCACTGATAGCAAAAATCTGGGAGGC 2940 Db

2941 AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000 QY
2941 AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000 Db
3001 TTGATTTCAATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060 QY
3001 TTGATTTCAATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060 Db
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGAGTCAATTTCAAAAGTGC 3120 QY
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGAGTCAATTTCAAAAGTGC 3120 Db
3121 GCCATTTCTGAGGATATGAGGTTTGCCTTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180 QY
3121 GCCATTTCTGAGGATATGAGGTTTGCCTTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180 Db
3181 AGCCACTGAATATATCTCAAGTCTTGTGATGAAGTTGATACAGATCAATCTGGTGTCTGT 3240 QY
3181 AGCCACTGAATATATCTCAAGTCTTGTGATGAAGTTGATACAGATCAATCTGGTGTCTGT 3240 Db
3241 CTGACAGAGAAATCCGAACACTGCTGCTACCAAGATTTCAAGAACTGCGGTTTGCAGG 3300 QY
3241 CTGACAGAGAAATCCGAACACTGCTGCTACCAAGATTTCAAGAACTGCGGTTTGCAGG 3300 Db
3301 ATTTGACAGGCTGGAAACATGCTGCTAATAATTTGCTCAAAATGCTTCTGCTGATATCA 3360 QY
3301 ATTTGACAGGCTGGAAACATGCTGCTAATAATTTGCTCAAAATGCTTCTGCTGATATCA 3360 Db
3361 CGCAGCTAAATTAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420 QY
3361 CGCAGCTAAATTAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420 Db
3421 TCACATAAAGTCTAGTAACAAACTGTAAACCCAGTAACTGATGATCCCAACCTGCCACCGG 3480 QY
3421 TCACATAAAGTCTAGTAACAAACTGTAAACCCAGTAACTGATGATCCCAACCTGCCACCGG 3480 Db
3481 AGGAAAACAAATATAGGTTTGAATCATGGAGAAAGAAATCGCTTTTAAATATGA 3540 QY
3481 AGGAAAACAAATATAGGTTTGAATCATGGAGAAAGAAATCGCTTTTAAATATGA 3540 Db
3541 TTGCTACCAAGCTTTCTCATGTTGGCCAGTTGGATGATGATGATGATGATGATGATGATG 3600 QY
3541 TTGCTACCAAGCTTTCTCATGTTGGCCAGTTGGATGATGATGATGATGATGATGATGATG 3600 Db
3601 AGTTTGTTCCTGAAATGACAACTTGACCAACATTAAGATGCTTCACAGAGTGAAGG 3660 QY
3601 AGTTTGTTCCTGAAATGACAACTTGACCAACATTAAGATGCTTCACAGAGTGAAGG 3660 Db
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAACTGCCAA 3720 QY
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAACTGCCAA 3720 Db
3721 GAGAGTATCGAAACGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATG 3780 QY
3721 GAGAGTATCGAAACGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATG 3780 Db
3781 ACAAAATGAAATTTGGACCCCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 3840 QY
3781 ACAAAATGAAATTTGGACCCCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 3840 Db
3841 CATTTTGTGAGCAGTTTAAATGCACTTAAAGCGGAAGATATTTCCAGAGAGGAGTAC 3900 QY
3841 CATTTTGTGAGCAGTTTAAATGCACTTAAAGCGGAAGATATTTCCAGAGAGGAGTAC 3900 Db
3901 ACAAGAGCTTAGTCCCAATCGAATCAGATGATGATGATGATGATGATGATGATGATGATG 3960 QY
3901 ACAAGAGCTTAGTCCCAATCGAATCAGATGATGATGATGATGATGATGATGATGATGATG 3960 Db
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTGTGATGATGAT 4020 QY
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTGTGATGATGAT 4020 Db
4021 GCTTAGCAGTTTGGCCCGGAGAGGAAATATCCAGTACCATGCTGTTTGTGTCGATGAA 4080 QY

4021	Db		GCATTAGCAGTTTGGCCGAAAGAGGAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4081
4081	Qy		TATAGCCCACTGACTAGGAAATATTTAAACCAACCACCTGAAAACCTTGTTGTCGAGCAGC	4140
4081	Db		TATAGCCCACTGACTAGGAAATATTTAAACCAACCACCTGAAAACCTTGTTGTCGAGCAGC	4140
4141	Qy		TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATAGGACCTGTCATCCTTTTATATAAAA	4200
4141	Db		TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATAGGACCTGTCATCCTTTTATATAAAA	4200
4201	Qy		GGCTCACTGACAAGAGACAGCTGTTAATTTCCCAACAGCAATCATTTGAGAGCTAACTTTAT	4260
4201	Db		GGCTCACTGACAAGAGACAGCTGTTAATTTCCCAACAGCAATCATTTGAGAGCTAACTTTAT	4260
4261	Qy		TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTTAAGAGGCTCCAGTCTTTGCAATCCAA	4320
4261	Db		TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTTAAGAGGCTCCAGTCTTTGCAATCCAA	4320
4321	Qy		AGCCTTTTGCTAAAGTTTTCACATTTTTTTTTTTTCATTTCCCATTTTTTAAGTAGTACTA	4380
4321	Db		AGCCTTTTGCTAAAGTTTTCACATTTTTTTTTTTTCATTTCCCATTTTTTAAGTAGTACTA	4380
4381	Qy		AGTTAACTAGTTATCTTTGCTTCGAGTATAACGAATTTGGATGCTTAAACCTATTTTTA	4440
4381	Db		AGTTAACTAGTTATCTTTGCTTCGAGTATAACGAATTTGGATGCTTAAACCTATTTTTA	4440
4441	Qy		TAGATGTTATTAANATATCGACAAATCACCTCTTATTGACAAATACCTAARATATGAG	4500
4441	Db		TAGATGTTATTAANATATCGACAAATCACCTCTTATTGACAAATACCTAARATATGAG	4500
4501	Qy		TTTTTATTAATTTTAAAGACTGTAATGGTCTTAAACCACCTAACTACTGAAAGAGCTCAATG	4560
4501	Db		TTTTTATTAATTTTAAAGACTGTAATGGTCTTAAACCACCTAACTACTGAAAGAGCTCAATG	4560
4561	Qy		ATTGACATCTGAATGCTTTGTAATTTAGACTTTAGCTTTAGCCCTTAAGAATGCTATGATTC	4620
4561	Db		ATTGACATCTGAATGCTTTGTAATTTAGACTTTAGCTTTAGCCCTTAAGAATGCTATGATTC	4620
4621	Qy		CGTGCAAGTCTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATATGTTT	4680
4621	Db		CGTGCAAGTCTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATATGTTT	4680
4681	Qy		TGGAATATGATCATATTTCAACAGAAAGTGCCTCATTTTAGAAATAGAGTACTGATGCGCA	4740
4681	Db		TGGAATATGATCATATTTCAACAGAAAGTGCCTCATTTTAGAAATAGAGTACTGATGCGCA	4740
4741	Qy		CTGCGCAATACAGTGTGCTTTGTTTAAATACTCATTTGGTATATTCAGTAGTACTCTCT	4800
4741	Db		CTGCGCAATACAGTGTGCTTTGTTTAAATACTCATTTGGTATATTCAGTAGTACTCTCT	4800
4801	Qy		CTCAGTTGGTTTTTGTATAGAACAGAGCCAGCAAACTTCTTTGTAAAAGGCTGGTTAGT	4860
4801	Db		CTCAGTTGGTTTTTGTATAGAACAGAGCCAGCAAACTTCTTTGTAAAAGGCTGGTTAGT	4860
4861	Qy		AAATATTTGAGGGCCACTGTGTCCTTTGTCATATACATTTCTTCTGCTGTTGTAGTTTGT	4920
4861	Db		AAATATTTGAGGGCCACTGTGTCCTTTGTCATATACATTTCTTCTGCTGTTGTAGTTTGT	4920
4921	Qy		TTTTTTTTCAAAACAACCTCTAAAATATGTAAAAACCATGTTTAGCTTCGAGCTGTACAAA	4980
4921	Db		TTTTTTTTCAAAACAACCTCTAAAATATGTAAAAACCATGTTTAGCTTCGAGCTGTACAAA	4980
4981	Qy		ACTGCCCAACAGCCAGATGTGACCCCTCAGGCCATCATTTGCGCAATCATCTGAGAAATATTT	5040
4981	Db		ACTGCCCAACAGCCAGATGTGACCCCTCAGGCCATCATTTGCGCAATCATCTGAGAAATATTT	5040
5041	Qy		TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTTGTCGCCAGGCTCGAG	5100
5041	Db		TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTCTGTTGTCGCCAGGCTCGAG	5100
5101	Qy		TGCAGTGGCGCAATCTCAGCTCACTGCAACTCCGCTCCGGGTTCAAGCAGTCTGTCTC	5160

Db	Seq	Score	DB	Length	Query Match	Best Local Similarity	Pred. No. 0;	Mismatches	Indels	Gaps	0;
5101	TGCAGTGGCGAATCTCAGCTCACTGCAACCTCCGCGCTCCCGGGTTCAAGCAGTCTCTGTC	5166									
5161	TCAGCCCTTCTGAGTAGCTGGGACTACAGGTGCAATGCCACACACCCCTGCTAAATTTTGTGA	5220									
5161	TCAGCCCTTCTGAGTAGCTGGGACTACAGGTGCAATGCCACACACCCCTGCTAAATTTTGTGA	5220									
5221	TTTTTGTAGTAGAGCGGGGGTTCCACCATATTTGGTTCAGGCTTATCTTGAACCTCCGACCTC	5280									
5221	TTTTTGTAGTAGAGCGGGGGTTCCACCATATTTGGTTCAGGCTTATCTTGAACCTCCGACCTC	5280									
5281	AGGTGATCCACCTGCTCTGCTCCCTCCCAAAAGTGTGAGATTACAGGCATAAGCCAGTGCAC	5340									
5281	AGGTGATCCACCTGCTCTGCTCCCTCCCAAAAGTGTGAGATTACAGGCATAAGCCAGTGCAC	5340									
5341	CCAGCCGAGAAATAGTATATTTTATGTATGTTGTTAAACCTTTGGCGCTTAGCCATATTTTATG	5400									
5341	CCAGCCGAGAAATAGTATATTTTATGTATGTTGTTAAACCTTTGGCGCTTAGCCATATTTTATG	5400									
5401	TCATAATACAAATGGATTGTGAAGAGCAGAAATCCATCAGTAACCTCTGACAGGTATTTTATG	5460									
5401	TCATAATACAAATGGATTGTGAAGAGCAGAAATCCATCAGTAACCTCTGACAGGTATTTTATG	5460									
5461	ATCATGATCTCAACATATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTTGAA	5520									
5461	ATCATGATCTCAACATATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTTGAA	5520									
5521	ATGTAATAATCTGTGTTTGTGCTGCTGAAGAGTTGTTGTTTCAAAAACCTGAAATCTCATAAA	5580									
5521	ATGTAATAATCTGTGTTTGTGCTGCTGAAGAGTTGTTGTTTCAAAAACCTGAAATCTCATAAA	5580									
5581	AAGTTAAATTTTGAATA 5597										
5581	AAGTTAAATTTTGAATA 5597										
RESULT 9											
US-09-895-072-20											
; Sequence 20, Application US/09895072											
; Patent No. US2002025550A1											
; GENERAL INFORMATION:											
; APPLICANT: CANFIELD, WILLIAM M											
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HY											
; FILE REFERENCE: 210119USOCONT											
; CURRENT APPLICATION NUMBER: US/09/895,072											
; CURRENT FILING DATE: 2001-07-02											
; PRIOR APPLICATION NUMBER: 60/153,831											
; PRIOR FILING DATE: 1999-09-14											
; PRIOR APPLICATION NUMBER: US 09/635,872											
; PRIOR FILING DATE: 2000-08-10											
; NUMBER OF SEQ ID NOS: 52											
; SOFTWARE: PatentIn version 3.1											
; SEQ ID NO 20											
; LENGTH: 3783											
; TYPE: DNA											
; ORGANISM: Homo sapiens											

Db 2347 GTGCTGAAAGATTGCAGAGTTGACCTTTTCCTGAGTGAGTGTAAGTGAATGGTTCAT 2406
QY 2565 GACCAGGGTCAGAAATCCACCCCTCGACTTGGAGACCACAGCAAGATTAGAGTGGAAACT 2624
Db 2407 GACCAGGGTCAGAAATCCACCCCTCGACTTGGAGACCACAGCAAGATTAGAGTGGAAACT 2466
QY 2625 CACACCCAAAACCATAGGCGGAAATGTGACAAAGAAAAGCCCCCATCTCTGATGTTT 2684
Db 2467 CACACCCAAAACCATAGGCGGAAATGTGACAAAGAAAAGCCCCCATCTCTGATGTTT 2526
QY 2685 CCACCTGGAAGCCAGATGACAAAAGAAAAGAAAATACACAGGGAAGAAAAGAGAACAGT 2744
Db 2527 CCACCTGGAAGCCAGATGACAAAAGAAAAGAAAATACACAGGGAAGAAAAGAGAACAGT 2586
QY 2745 AGAATGAGGAAAATGCTGAAAATTCATAGGCGTTACTGAAGTGTGTACTTGGAAAGAAAG 2804
Db 2587 AGAATGAGGAAAATGCTGAAAATTCATAGGCGTTACTGAAGTGTGTACTTGGAAAGAAAG 2646
QY 2805 CTGACGATATACAGATAGTTACTTGGGCTTTTTCATCGGAGGAGAAAAGAAAGTATTTT 2864
Db 2647 CTGACGATATACAGATAGTTACTTGGGCTTTTTCATCGGAGGAGAAAAGAAAGTATTTT 2706
QY 2865 CAAGATCTTCTCGAGGAAGAGTCAATGGAAGACAAATTTGGCATACTTCACTGATAGC 2924
Db 2707 CTAGATCTTCTCGAGGAAGAGTCAATGGAAGACAAATTTGGCATACTTCACTGATAGC 2766
QY 2925 AAAAATACTGGAGGGAACCTAAAGATACATTTGCAGATTCCTCTCAGATGTAAATAAA 2984
Db 2767 AAGAATACTGGAGGGAACCTAAAGATACATTTGCAGATTCCTCTCAGATGTAAATAAA 2826
QY 2985 ATTCTAAATAGCAAGTTTGGATTCCATCGGGAAGTCCCTGCTCAGATGCCCTCACATG 3044
Db 2827 ATTCTAAATAGCAAGTTTGGATTCCATCGGGAAGTCCCTGCTCAGATGCCCTCACATG 2886
QY 3045 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAGC 3104
Db 2887 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAGC 2946
QY 3105 TCATTTCACAAAGTGGCCATTCGAGGATATGAGTTTGCCTTCTCTTATTTTATAT 3164
Db 2947 TCATTTCACAAAGTGGCCATTCGAGGATATGAGTTTGCCTTCTCTTATTTTATAT 3006
QY 3165 CTCATGATGTCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224
Db 3007 CTCATGATGTCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066
QY 3225 CAATCTGGTGTCTGTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCAGAACTG 3284
Db 3067 CAATCTGGTGTCTGTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCAGAACTG 3126
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGGTTCTGGAACACATGCTTAATAAATTTGCTCAAAAATG 3344
Db 3127 CCGTTAAGTTTGCAGGATTTGACAGGTTCTGGAACACATGCTTAATAAATTTGCTCAAAAATG 3186
QY 3345 CTTCTGCTGATATCAGCGAGCTAAATAATATTCACCAACTCAGGAAATCCTACTATGAT 3404
Db 3187 CTTCTGCTGATATCAGCGAGCTAAATAATATTCACCAACTCAGGAAATCCTACTATGAT 3246
QY 3405 CCCAACCTGCCACCGGTCACTAAAGTCTAGTACAAACTGTAAACCAAGTAACTGCACAA 3464
Db 3247 CCCAACCTGCCACCGGTCACTAAAGTCTAGTACAAACTGTAAACCAAGTAACTGCACAA 3306
QY 3465 ATCCACAAAGCATATAAGGACAAAACAAATATATAGTTTGAATTCATGGGAGAGAGAA 3524
Db 3307 ATCCACAAAGCATATAAGGACAAAACAAATATATAGTTTGAATTCATGGGAGAGAGAA 3366
QY 3525 ATCCGTTTTAAATGATTCGTACCAAGTTTCTCATGTGTTGGTGGCCAGTTGGATGACATA 3584
Db 3367 ATCCGTTTTAAATGATTCGTACCAAGTTTCTCATGTGTTGGTGGCCAGTTGGATGACATA 3426
QY 3585 AGAAAAACCCCTAGGAAGTTTGTGTTGCTGTAATGACAAACATTTGACCAACAATTCATAAGAT 3644
Db 3427 AGAAAAACCCCTAGGAAGTTTGTGTTGCTGTAATGACAAACATTTGACCAACAATTCATAAGAT 3486

QY 3645 GCTCAGACAGTGAAGCTGTTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3704
Db 3487 GCTCAGACAGTGAAGCTGTTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3546
QY 3705 CAATTGTAAGTCCCAAGAGAGTATCGAAACCGTTTCTTCTATGATCATGATGATGAGGAA 3764
Db 3547 CAATTGTAAGTCCCAAGAGAGTATCGAAACCGTTTCTTCTATGATCATGATGATGAGGAA 3606
QY 3765 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3824
Db 3607 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3666
QY 3825 ATGTTTACTATATCTCTATTTTCTGAGAGTTAATTTGACCTTAAAGCGGAGATATTT 3884
Db 3667 ATGTTTACTATATCTCTATTTTCTGAGCAGTTAATTTGCACCTTAAAGCGGAGATATTT 3726
QY 3885 CCCAGAAGGAGATACACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAGATC 3941
Db 3727 CCCAGAAGGAGATACACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAGATC 3783

RESULT 10
US-09-986-552-20
; Sequence 20, Application US/09986552
; Patent No. US20030150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE;
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-986-552-20

Query Match 67.2%; Score 3761; DB 9; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 165 ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCCTGCCACAGGTATGGGCTC 224
Db 7 ATGGGGTTCAAGCTCTTGACAGAGACAAACCTATACCTGCCTGCCACAGGTATGGGCTC 66
QY 225 TACGTGTGCTTCTTGGGGTCTGTTGTCACCATCGTCTCGGCTTCCAGTTCGGAGAGTG 284
Db 67 TACGTGTGCTTCTTGGGGTCTGTTGTCACCATCGTCTCGGCTTCCAGTTCGGAGAGTG 126
QY 285 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGACAAATATT 344
Db 127 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGACAAATATT 186
QY 345 GCTGGAAGTCTTTTCAGATCGGCTTGTCTGCCCATCGCGATTGACGTTGTTTACACC 404
Db 187 GCTGGAAGTCTTTTCAGATCGGCTTGTCTGCCCATCGCGATTGACGTTGTTTACACC 246
QY 405 TGGGTCAATGGCACAGATCTTGAACTACTGAAGGAATACACAGAGTTCAGAGAACAGATG 464
Db 247 TGGGTCAATGGCACAGATCTTGAACTACTGAAGGAATACACAGAGTTCAGAGAACAGATG 306
QY 465 GAGGAGGAGCAGAAAGCAATGAGAAATCCTTTGGGAAAAACACAAACCGAACCTACTAAG 524
Db 307 GAGGAGGAGCAGAAAGCAATGAGAAATCCTTTGGGAAAAACACAAACCGAACCTACTAAG 366
QY 525 AAGAGTGGAAGCAGTTAGAGTGTGTTGTTTAAACACACTGCAATTAAGTTCGCAATGTTCTGTA 584

367	Db		AAGAGT	GAGAAG	CAGT	TTAGAGT	GTGTTG	CTAACA	CAC	TGCAT	TAAAGT	GTGCCAA	TGCTTG	TC	426			
585	QY		CTGAC	CCAGC	CGCTG	CCAGC	CAACAT	CAC	CCTG	AAGG	ACGTG	GC	CACTCT	TTTAT	CTCTTC	644		
427	Db		CTGAC	CCAGC	CGCTG	CCAGC	CAACAT	CAC	CCTG	AAGG	ACGTG	GC	CACTCT	TTTAT	CTCTTC	486		
645	QY		TTTCAT	TCTG	CCAGT	GCACAT	TTTCAAT	TGTTG	CAAAAC	CAAAAA	CCCTTCT	CTAC	CAATGTC			704		
487	Db		TTTCAT	TCTG	CCAGT	GCACAT	TTTCAAT	TGTTG	CAAAAC	CAAAAA	CCCTTCT	CTAC	CAATGTC			546		
705	QY		TCAGT	TGTTG	TTTGTG	ACAGT	ACTA	AGAGT	GTG	AAGAT	GC	CCCACTCT	CG	AGCTG	CTTAAA	764		
547	Db		TCAGT	TGTTG	TTTGTG	ACAGT	ACTA	AGAGT	GTG	AAGAT	GC	CCCACTCT	CG	AGCTG	CTTAAA	606		
765	QY		GGAAT	TAG	CAC	A	GACAGAT	GAGGGG	TAC	TGTG	CAACAG	AATAAG	AGTCC	CTGGA		824		
607	Db		GGAAT	TAG	CAC	A	GACAGAT	GAGGGG	TAC	TGTG	CAACAG	AATAAG	AGTCC	CTGGA		666		
825	QY		TTAGT	GCTA	ATG	CAAGAT	TTTGGCTT	CCTG	AGT	GGATTT	CCA	CCAA	CAT	TCAAG	GAACA	884		
667	Db		TTAGT	GCTA	ATG	CAAGAT	TTTGGCTT	CCTG	AGT	GGATTT	CCA	CCAA	CAT	TCAAG	GAACA	726		
885	QY		AATCAA	CT	AAAA	CAAAAT	TCCAG	AAAAAT	CTTTC	CTTAA	AGT	CAAACT	TGTTG	CACTG		944		
727	Db		AATCAA	CT	AAAA	CAAAAT	TCCAG	AAAAAT	CTTTC	CTTAA	AGT	CAAACT	TGTTG	CACTG		786		
945	QY		TATTC	AG	AGCC	AGTGT	AG	CGCTTCT	TA	AAAT	TAACCC	CAAG	ATTTT	CA	AGAAATG	1004		
787	Db		TATTC	AG	AGCC	AGTGT	AG	CGCTTCT	TA	AAAT	TAACCC	CAAG	ATTTT	CA	AGAAATG	846		
1005	QY		AAT	TAAG	CAAACT	TA	AG	AGCA	CTG	ATG	AGAA	AGAACT	GA	CTG	CACTG	1064		
847	Db		AAT	TAAG	CAAACT	TA	AG	AGCA	CTG	ATG	AGAA	AGAACT	GA	CTG	CACTG	906		
1065	QY		TATTTAT	TATGG	ATCT	GAG	CGCCAT	CAG	CCAGTCT	TAAG	CAGGAT	GAAG	ACATCT	CTG	CC	1124		
907	Db		TATTTAT	TATGG	ATCT	GAG	CGCCAT	CAG	CCAGTCT	TAAG	CAGGAT	GAAG	ACATCT	CTG	CC	966		
1125	QY		AGTCG	TTTTG	NAG	ATAC	AG	AGAACT	CAGGT	ACTCAT	TGGAT	CTAT	CG	AGAG	GCATGCA	1184		
967	Db		AGTCG	TTTTG	NAG	ATAC	AG	AGAACT	CAGGT	ACTCAT	TGGAT	CTAT	CG	AGAG	GCATGCA	1026		
1185	QY		CCATG	GGTTC	GG	AAATATTTT	CAT	TGTCA	CAAA	CGGG	CAGATTC	CACT	CGCT	GGCT	GAAC	1244		
1027	Db		CCATG	GGTTC	GG	AAATATTTT	CAT	TGTCA	CAAA	CGGG	CAGATTC	CACT	CGCT	GGCT	GAAC	1086		
1245	QY		GAC	AATCCT	CG	AGTGA	CAAT	TAGT	AA	CA	CCAG	ATGTTTT	TG	CAAA	TTTGA	GC	1304	
1087	Db		GAC	AATCCT	CG	AGTGA	CAAT	TAGT	AA	CA	CCAG	ATGTTTT	TG	CAAA	TTTGA	GC	1146	
1305	QY		CCT	ACCTTT	TAG	TTCA	CCCTGCT	TAT	TGA	AGTCA	CACTCAT	CG	ATCG	AGAG	GGCTG	TCCAG	1364	
1147	Db		CCT	ACCTTT	TAG	TTCA	CCCTGCT	TAT	TGA	AGTCA	CACTCAT	CG	ATCG	AGAG	GGCTG	TCCAG	1206	
1365	QY		AAG	TTTATTT	ACCT	TA	ATATGAT	GTG	TCAT	GT	TGG	GAAG	ATG	TCT	CGCC	CAGAT	1424	
1207	Db		AAG	TTTATTT	ACCT	TA	ATATGAT	GTG	TCAT	GT	TGG	GAAG	ATG	TCT	CGCC	CAGAT	1266	
1425	QY		TAC	AGTCA	CTCC	AAAG	CC	CAGAG	GGTTT	TAT	TG	CAT	GGCTG	CG	CAAACT	GTG	CCGAG	1484
1267	Db		TAC	AGTCA	CTCC	AAAG	CC	CAGAG	GGTTT	TAT	TG	CAT	GGCTG	CG	CAAACT	GTG	CCGAG	1326
1485	QY		GGCT	GCC	CAG	GTTCCT	CG	ATTTA	AG	AGT	CA	CCCTTCT	CT	GAAG	GGT	TAAT	1544	
1327	Db		GGCT	GCC	CAG	GTTCCT	CG	ATTTA	AG	AGT	CA	CCCTTCT	CT	GAAG	GGT	TAAT	1386	
1545	QY		TGCG	AT	TGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	1604	
1387	Db		TGCG	AT	TGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	TGGG	AT	1446	
1505	QY		GGT	GG	AG	GT	AT	TGG	AG	GT	TGG	AG	GT	TGG	AG	GT	1664	

1447	GGTGGAGGTA	CTGGGAGTA	TATGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAC	1509
1665	AGTGTCCTTACT	GTAAATCAGGAGTGGGAAAT	TCTGGCTCGCTGATAAGGTTCTGTGAC	1724
1507	AGTGTCCTTACT	GTAAATCAGGAGTGGGAAAT	TCTGGCTCGCTGATAAGGTTCTGTGAC	1566
1725	CAAGCATGCAATG	CTTGTCCTGTGGGTTTGATGCTGGGCACTGTGGCAGATCATTTT	1784	
1567	CAAGCATGCAATG	CTTGTCCTGTGGGTTTGATGCTGGGCACTGTGGCAGATCATTTT	1626	
1785	CATGAATTTGTA	TAAAGTGATCCCTTCTCCCAAACAGAGCTCACTATATTTATTTCCAAAAGGT	1844	
1627	CATGAATTTGTA	TAAAGTGATCCCTTCTCCCAAACAGAGCTCACTATATTTATTTCCAAAAGGT	1686	
1845	GAATCCCTGCCTT	ATTTTTCAGCTTTGCAGAGTAGCCAAAGAGGAGTTGAAGGTGCCTAT	1904	
1687	GAATCCCTGCCTT	ATTTTTCAGCTTTGCAGAGTAGCCAAAGAGGAGTTGAAGGTGCCTAT	1746	
1905	AGTGCAATCAAT	TAAATTTGCAGATGCTTCTATTGGCCACAGGTGGAACCATCAACCTC	1964	
1747	AGTGCAATCAAT	TAAATTTGCAGATGCTTCTATTGGCCACAGGTGGAACCATCAACCTC	1806	
1965	ATAATGCAAGTGA	ATGAATGCCACCAATACTTTTAATCTCACGTTTCAAAATACA	2024	
1807	ATAATGCAAGTGA	ATGAATGCCACCAATACTTTTAATCTCACGTTTCAAAATACA	1866	
2025	AACGATGAAGAG	TTCAAAATGTCAGATAACAGTGGAGTGGACACAGGAGGAGCCAAA	2084	
1867	AACGATGAAGAG	TTCAAAATGTCAGATAACAGTGGAGTGGACACAGGAGGAGCCAAA	1926	
2085	CTGAATTTCTAC	GGCCCAAGGGTTACGAAAATTTAGTTAGTTCCTCAATCAACACTTCTTCCA	2144	
1927	CTGAATTTCTAC	GGCCCAAGGGTTACGAAAATTTAGTTAGTTCCTCAATCAACACTTCTTCCA	1986	
2145	GAGCGGAAATTC	CTTTTGTAGGATATTTCCCAAGAAAACCGTCTCCGAACTTTAAGAGA	2204	
1987	GAGCGGAAATTC	CTTTTGTAGGATATTTCCCAAGAAAACCGTCTCCGAACTTTAAGAGA	2046	
2205	CATGATGTTAACT	CAACAGGAGAGCCAGGAGGTGAAAATTTCCCTCGTAAAATATT	2264	
2047	CATGATGTTAACT	CAACAGGAGAGCCAGGAGGTGAAAATTTCCCTCGTAAAATATT	2106	
2265	TCACTCTCTTCA	AAAGAGCCCAAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACAT	2324	
2107	TCACTCTCTTCA	AAAGAGCCCAAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACAT	2166	
2325	GGAGCATCACTTT	GAAAGATACAAATTTGTCCTCAAGTCAGCTTCTGAGATCAATTTCTG	2384	
2167	GGAGCATCACTTT	GAAAGATACAAATTTGTCCTCAAGTCAGCTTCTGAGATCAATTTCTG	2226	
2385	ATGAATCTCAC	AGCATGTCAAAATTAATAACAGATGAACCAAAATGC	2444	
2227	ATGAATCTCAC	AGCATGTCAAAATTAATAACAGATGAACCAAAATGC	2386	
2445	AGTTTGTGGCTC	CAACAGGAAAACAGGTTTCATAAAGCATCTTCCGCAACAGCTTAGGA	2504	
2287	AGTTTGTGGCTC	CAACAGGAAAACAGGTTTCATAAAGCATCTTCCGCAACAGCTTAGGA	2346	
2505	GTGCTGAAACAT	TGCAAGAGTTGACTTTTCTGCAAGTGAAGAGTGAAGTGGTGCAT	2564	
2347	GTGCTGAAACAT	TGCAAGAGTTGACTTTTCTGCAAGTGAAGAGTGAAGTGGTGCAT	2406	
2565	GACAGGTCAGAA	ATCCACCCCTTGGACTTTGGAGACCAAGCAAGATTTTAGAGTAGGAACT	2624	
2407	GACAGGTCAGAA	ATCCACCCCTTGGACTTTGGAGACCAAGCAAGATTTTAGAGTAGGAACT	2466	
2625	CACACCCAAAAC	CAATAGCGGGAATGTGACAAAAGAAAAGCCGCCCATCTCTGATTTGTT	2684	
2467	CACACCCAAAAC	CAATAGCGGGAATGTGACAAAAGAAAAGCCGCCCATCTCTGATTTGTT	2526	
2685	CCACTGGAAAGC	CCAGATGACAAAGAAAAGAAAATCAACAGGGAAGAAAAGAGAACAGT	2744	
2527	CCACTGGAAAGC	CCAGATGACAAAGAAAAGAAAATCAACAGGGAAGAAAAGAGAACAGT	2586	

QY	2745	AGAAATGGAGGAAATGCTGAAATTCACATAGGCGCTTACTGAAAGTGTTACTTTGGAAGAAG	2804
Db	2587	AGAAATGGAGGAAATGCTGAAATTCACATAGGCGCTTACTGAAAGTGTTACTTTGGAAGAAG	2646
QY	2805	CTGCAGCATTACACAGATAGTACTTGGCGCTTTTGGCATGGGAGAAAAAAGATATTTC	2864
Db	2647	CTGCAGCATTACACAGATAGTACTTGGCGCTTTTGGCATGGGAGAAAAAAGATATTTC	2706
QY	2865	CAAGATCTTCTCGAGGAGAGAGTGTCATTGAGACACAAATGGGCATCTTCACGTGATAGC	2924
Db	2707	CTAGATCTTCTCGAGGAGAGAGTGTCATTGAGACACAAATGGGCATCTTCACGTGATAGC	2766
QY	2925	AAAAATACTGGGAGGCAACTAAAAAGATACATTTTCAGATTCCCTCAGATATGTAATAAAA	2984
Db	2767	AAGAAATACTGGGAGGCAACTAAAAAGATACATTTTCAGATTCCCTCAGATATGTAATAAAA	2826
QY	2985	ATTCTAAATAGCAGATTGTGNTTCACATCGCGGAAAGTCCCTGCTCACAATGCCCTCACATG	3044
Db	2827	ATTCTAAATAGCAGATTGTGNTTCACATCGCGGAAAGTCCCTGCTCACAATGCCCTCACATG	2886
QY	3045	ATTGACCGGATGTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACG	3104
Db	2887	ATTGACCGGATGTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACG	2946
QY	3105	TCATTTCAAAAGTGCGCATCTCGAGATATGAGTTTGCCTTCTCTTATTTTATAT	3164
Db	2947	TCATTTCAAAAGTGCGCATCTCGAGATATGAGTTTGCCTTCTCTTATTTTATAT	3006
QY	3165	CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT	3224
Db	3007	CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT	3066
QY	3225	CAATCTCGTGTCTTGCTGACAGAGAAATCCGAACATCGGCTACCGAAATTCACGAACCTG	3284
Db	3067	CAATCTCGTGTCTTGCTGACAGAGAAATCCGAACATCGGCTACCGAAATTCACGAACCTG	3126
QY	3285	CCGTTAAGTTTGACAGGATTTGACAGGTCCTGGAACACATGCTTAATAAATGCTCCTCAAAATG	3344
Db	3127	CCGTTAAGTTTGACAGGATTTGACAGGTCCTGGAACACATGCTTAATAAATGCTCCTCAAAATG	3186
QY	3345	CTTCCTGCTGATATCACGCACTAAATAATATCCCACTCAGGAATCCTACTATGAT	3404
Db	3187	CTTCCTGCTGATATCACGCACTAAATAATATCCCACTCAGGAATCCTACTATGAT	3246
QY	3405	CCCAACCTGCCACCGGTCACTTAAAGTCTAGTACAAACCTGTAACACAGTAACCTGACAAA	3464
Db	3247	CCCAACCTGCCACCGGTCACTTAAAGTCTAGTACAAACCTGTAACACAGTAACCTGACAAA	3306
QY	3465	ATCCACAAAGCATATAAGGCAAAAAAATAATATAGGTTTGAAATCATGGGAGAGAAAGAA	3524
Db	3307	ATCCACAAAGCATATAAGGCAAAAAAATAATATAGGTTTGAAATCATGGGAGAGAAAGAA	3366
QY	3525	ATCGCTTTTAAATGATTCGTACCAACGGTTCTCATGTGTTGCCAGTTGGATGACATA	3584
Db	3367	ATCGCTTTTAAATGATTCGTACCAACGGTTCTCATGTGTTGCCAGTTGGATGACATA	3426
QY	3585	AGAAAAACCCCTAGGAAAGTTGTTGGCTGAAATGACAAACATGACCAACAATCATAAAGAT	3644
Db	3427	AGAAAAACCCCTAGGAAAGTTGTTGGCTGAAATGACAAACATGACCAACAATCATAAAGAT	3486
QY	3645	GCTCAGACAGTGAAGGCTGTTCTCAGGCACTTCTATGAATCCATGTTCCCGATCCTTCC	3704
Db	3487	GCTCAGACAGTGAAGGCTGTTCTCAGGCACTTCTATGAATCCATGTTCCCGATCCTTCC	3546
QY	3705	CAATTTGAACTGCCAAGAGATATCGAAACCGTTTCTTATATGCTATGAGCTGACAGGAA	3764
Db	3547	CAATTTGAACTGCCAAGAGATATCGAAACCGTTTCTTATATGCTATGAGCTGACAGGAA	3606
QY	3765	TGGAGGCTTATCGAGACAAATGGAAGTTTTGGACCCATTTGTGTACTAGCAACATTCGAT	3824
Db	3607	TGGAGGCTTATCGAGACAAATGGAAGTTTTGGACCCATTTGTGTACTAGCAACATTCGAT	3666

Qy	3825	ATGTTTACTATATTCATTCTTTTGCTGGACAGTAAATGCACCTTAAGCGGAAGATAATTT	3884
Db	3667	ATGTTTACTATATTCATTCTTTTGCTGGACAGTAAATGCACCTTAAGCGGAAGATAATTT	3726
Qy	3885	CCCAGAAGGAGGATACACAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC	3941
Db	3727	CCCAGAAGGAGGATACACAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC	3783
 RESULT 11 US-10-306-686-20 ; Sequence 26, Application US/10306686 ; Publication No. US20030148460A1 ; GENERAL INFORMATION: ; APPLICANT: CANFIELD, WILLIAM ; TITLE OF INVENTION: PHOSPHOESTER ALPHA-CLONASE OF THE LYSOSOMAL TARGETING ; FILE REFERENCE: 230397US77DIV ; CURRENT APPLICATION NUMBER: US/10/306,686 ; CURRENT FILING DATE: 2002-11-29 ; PRIOR APPLICATION NUMBER: 09/636,596 ; PRIOR FILING DATE: 2000-08-10 ; PRIOR APPLICATION NUMBER: 60/153,831 ; PRIOR FILING DATE: 1999-08-14 ; NUMBER OF SEQ ID NOS: 52 ; SOFTWARE: PatentIn version 3.1 ; SEQ ID NO 20 ; LENGTH: 3783 ; TYPE: DNA ; ORGANISM: Homo sapiens US-10-306-686-20			
 Query Match 67.2%; Score 3761; DB 15; Length 3783; Best Local Similarity 99.7%; Pred. No. 0; Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			
Qy	165	ATGCTGTTCAGCTCCTGCGAGACAAAACCTATACCTGCTCCTGTCGCCACAGGTATGGGCTC	224
Db	7	ATGGGTTCAAGCTCTTTCGAGAGACAAACCTATACCTGCTCCTGTCGCCACAGGTATGGGCTC	66
Qy	225	TAGTGTGCTCTTTGGCGCTGCTGTGTCAACATCGTCTCGCCCTTCAGTTCGGAGAGGTG	284
Db	67	TAGTGTGCTCTTTGGCGCTGCTGTGTCAACATCGTCTCGCCCTTCAGTTCGGAGAGGTG	126
Qy	285	GTTCTGGAATGGAGCGAGATCAATACCATTGTTTTGTTGATTCCTATAGAGACAATATT	344
Db	127	GTTCCTGGAATGGAGCGAGATCAATACCATTGTTTTGTTGATTCCTATAGAGACAATATT	186
Qy	345	GCTGGAAAGTCCCTTCAGAAATCGGCTTTGTCGCCCATGCGGATTCGAGTGTGTTTACACC	404
Db	187	GCTGGAAAGTCCCTTCAGAAATCGGCTTTGTCGCCCATGCGGATTCGAGTGTGTTTACACC	246
Qy	405	TGGTGTGATGGCACAGATCTTGAACCTACTGAAGGAACTACAGCAGGTTCAGAGAACAGATG	464
Db	247	TGGTGTGATGGCACAGATCTTGAACCTACTGAAGGAACTACAGCAGGTTCAGAGAACAGATG	306
Qy	465	GAGGAGGAGCAGAAAGCAATAGAGAAAACTCTTGGGAAAAACAACACGGAACTACTTAAG	524
Db	307	GAGGAGGAGCAGAAAGCAATAGAGAAAACTCTTGGGAAAAACAACACGGAACTACTTAAG	366
Qy	525	AAGAGTGAGAGCAGTTAGAGTCTTTTGCTTAACACATGCTGATTAAGTGCCTGCTTGA	584
Db	367	AAGAGTGAGAGCAGTTAGAGTCTTTTGCTTAACACATGCTGATTAAGTGCCTGCTTGA	426
Qy	585	CTGACCCAGCCCTGCCAGCCAAACATCACCCCTGAAGGAGTGCCTCTCTTTATCCTCTCT	644
Db	427	CTGACCCAGCCCTGCCAGCCAAACATCACCCCTGAAGGAGTGCCTCTCTTTATCCTCTCT	486
Qy	645	TTTCATCTCGCAGTGACATTTTCAATGTGCAAAACCAAACCAAAACCCCTCTACCAATGTC	704
Db	487	TTTCATCTCGCAGTGACATTTTCAATGTGCAAAACCAAACCAAAACCCCTCTACCAATGTC	546
Qy	705	TCAGTTGTGTTTTTGACACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAA	764

Db 547 TCAGTTGTTGTTTGTGACAGTACTAAGGATGTTGAAGATGCCCACTCTCGACTGCTTAAA 606
Qy 765 GGAAATAGCAGACAGACATGATGGGGGGTACTTGACACAGATAAAGAACTCCCTCGA 824
Db 607 GGAAATAGCAGACAGACATGATGGGGGGTACTTGACACAGATAAAGAACTCCCTCGA 666
Qy 825 TTAGTGTCTAATGCAAGATTTGGCTTCTCGAGTGGATTTCCACCAACATTCAAGGAAACA 884
Db 667 TTAGTGTCTAATGCAAGATTTGGCTTCTCGAGTGGATTTCCACCAACATTCAAGGAAACA 726
Qy 885 AATCACTAATAAACAATAATGCGAGAAATCTTCTCTAAGTCAAACTGTTGAGTTG 944
Db 727 AATCACTAATAAACAATAATGCGAGAAATCTTCTCTAAGTCAAACTGTTGAGTTG 786
Qy 945 TATTGAGGCCAGTGTAGGGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 1004
Db 787 TATTGAGGCCAGTGTAGGGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 846
Qy 1005 AATAAGCAAACTAAGAAAGAAATGACATGACATGATGGAAGAAAGCTGACCAATAGTCTGCA 1064
Db 847 AATAAGCAAACTAAGAAAGAAATGACATGATGGAAGAAAGCTGACCAATAGTCTGCA 906
Qy 1065 TATTATTATGGGATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124
Db 907 TATTATTATGGGATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966
Qy 1125 AGTCGTTTTGAAGATRAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGCAATGCA 1184
Db 967 AGTCGTTTTGAAGATRAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGCAATGCA 1026
Qy 1185 CCATGGTTCGGAAATATTTTCATTGTGTCACCAAGGGCAGATTCATCTCGTGGTACCTT 1244
Db 1027 CCATGGTTCGGAAATATTTTCATTGTGTCACCAAGGGCAGATTCATCTCGTGGTACCTT 1086
Qy 1245 GACAACTCCTCGAGTGAACAATAGTAACACACAGAGATGTTTTGCAAAATTTGAGCCACTTG 1304
Db 1087 GACAACTCCTCGAGTGAACAATAGTAACACACAGAGATGTTTTGCAAAATTTGAGCCACTTG 1146
Qy 1305 CTTACCTTTAGTCACTGCTGTAATGAAAGTCAATTCATCGATGCAAGGGCTGTCGCCAG 1364
Db 1147 CTTACCTTTAGTCACTGCTGTAATGAAAGTCAATTCATCGATGCAAGGGCTGTCGCCAG 1206
Qy 1365 AAGTTTATTTACCTTAATGATGATGTCATGTTGGGAAGATGCTGGCCAGATGATTTT 1424
Db 1207 AAGTTTATTTACCTTAATGATGATGTCATGTTGGGAAGATGCTGGCCAGATGATTTT 1266
Qy 1425 TACAGTCACTCCAAAGCCAGAGAGTTTATTTGACATGGCTGTGCAAACTGTGCCGAG 1484
Db 1267 TACAGTCACTCCAAAGCCAGAGAGTTTATTTGACATGGCTGTGCAAACTGTGCCGAG 1326
Qy 1485 GGCTGCCAGGTTCTCGATTAAGATGCTATTTGTAAGAGCTTTGTAATTAATTCAGCC 1544
Db 1327 GGCTGCCAGGTTCTCGATTAAGATGCTATTTGTAAGAGCTTTGTAATTAATTCAGCC 1386
Qy 1545 TGCATTTGGGATGGTGGGATTTGCTCTGGAACAGTGGAGGAGTCTGCTATATTTGCAAG 1604
Db 1387 TGCATTTGGGATGGTGGGATTTGCTCTGGAACAGTGGAGGAGTCTGCTATATTTGCAAG 1446
Qy 1605 GGTGGAGGATCTGGAGTATGAGTTGGACACCCCTGGAGTTTGGTGGAGGATTAAC 1664
Db 1447 GGTGGAGGATCTGGAGTATGAGTTGGACACCCCTGGAGTTTGGTGGAGGATTAAC 1506
Qy 1665 AGTGTCTCTTACTGTAATCAGGATGTCGAAATCTCTGGCTCGCTGATTAAGTTCTGTGAC 1724
Db 1507 AGTGTCTCTTACTGTAATCAGGATGTCGAAATCTCTGGCTCGCTGATTAAGTTCTGTGAC 1566
Qy 1725 CAAGCATGCAATGTCCTGTGGGTTTGTGCTGGGCACTGTGGGCAAGATCATTTT 1784
Db 1567 CAAGCATGCAATGTCCTGTGGGTTTGTGCTGGGCACTGTGGGCAAGATCATTTT 1626
Qy 1785 CATGAATTTGTAATGATGCTCTTCTCCCAACACAGACTCACTATATTTTCCAAAGGT 1844
Db 1627 CATGAATTTGTAATGATGCTCTTCTCCCAACACAGACTCACTATATTTTCCAAAGGT 1686

Qy 1845 GAATGCTGCTCTTATTTTTCAGCTTTGAGAAAGTAGCCAAAAGAGAGGTTGAAGTGGCTAT 1904
Db 1687 GAATGCTGCTCTTATTTTTCAGCTTTGAGAAAGTAGCCAAAAGAGAGGTTGAAGTGGCTAT 1746
Qy 1905 AGTGACAAATCAATTAATTCGACATGCTTCTATTGCGCAACAAAGTGGAAACCAATCCACCTC 1964
Db 1747 AGTGACAAATCAATTAATTCGACATGCTTCTATTGCGCAACAAAGTGGAAACCAATCCACCTC 1806
Qy 1965 ATAATGACAGTGGAAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAAATACA 2024
Db 1807 ATAATGACAGTGGAAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAAATACA 1866
Qy 2025 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGTGACACAGGGGAGGACCAAAA 2084
Db 1867 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGTGACACAGGGGAGGACCAAAA 1926
Qy 2085 CTGAATTTCTAGCGCCCAAGAGGGTTACGAAAATTTAGTTAGTCCCAATACATCTTTTCCA 2144
Db 1927 CTGAATTTCTAGCGCCCAAGAGGGTTACGAAAATTTAGTTAGTCCCAATACATCTTTTCCA 1986
Qy 2145 GAGCGGAAATCTCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAAAGTTTAAGAGA 2204
Db 1987 GAGCGGAAATCTCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAAAGTTTAAGAGA 2046
Qy 2205 CATGATGTTAACTCAACAGAGAGCCCGAGAGAGGTGAATAATTCCTCGGTAAATATT 2264
Db 2047 CATGATGTTAACTCAACAGAGAGCCCGAGAGAGGTGAATAATTCCTCGGTAAATATT 2106
Qy 2265 TCATCTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAAAT 2324
Db 2107 TCATCTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAAAT 2166
Qy 2325 GGAGACATCACTTTGAAAGGATCAATTTGTCCAAAGTCCAGCTTGTGAGATCATTTCTG 2384
Db 2167 GGAGACATCACTTTGAAAGGATCAATTTGTCCAAAGTCCAGCTTGTGAGATCATTTCTG 2226
Qy 2385 ATGAACCTCAGACATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGAC 2444
Db 2227 ATGAACCTCAGACATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGAC 2286
Qy 2445 AGTTTGTGGCTCCACAGGAAAAACAGGTTCAATAAAGCATCTTGCACAAACAGCTTAGGA 2504
Db 2287 AGTTTGTGGCTCCACAGGAAAAACAGGTTCAATAAAGCATCTTGCACAAACAGCTTAGGA 2346
Qy 2505 GTGTCTGAAGATGACAGAGTTGACATTTTCCCTGAGTGAAGTGAAGTGAATGGTCAAT 2564
Db 2347 GTGTCTGAAGATGACAGAGTTGACATTTTCCCTGAGTGAAGTGAAGTGAATGGTCAAT 2406
Qy 2565 GACCAGGCTCAGAAATCCACCCCTGGACTTGGAGACCAACAGAGATTTAGAGTGGAAACT 2624
Db 2407 GACCAGGCTCAGAAATCCACCCCTGGACTTGGAGACCAACAGAGATTTAGAGTGGAAACT 2466
Qy 2625 CACACCCAAAAAACATAGCGGAAATCTGACAAAAGAAAAGCCCATCTCTGATTTGTT 2684
Db 2467 CACACCCAAAAAACATAGCGGAAATCTGACAAAAGAAAAGCCCATCTCTGATTTGTT 2526
Qy 2685 CCATGGAAAGCCAGATGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGT 2744
Db 2527 CCATGGAAAGCCAGATGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGT 2586
Qy 2745 AGAATGGAGGAAATGCTGAAAAATCAATAGGCGTTACTGAAGTGTACTTTGGAGAAAG 2804
Db 2587 AGAATGGAGGAAATGCTGAAAAATCAATAGGCGTTACTGAAGTGTACTTTGGAGAAAG 2646
Qy 2805 CTGAGACATTAACAGATAGTTACTTTGGCTTTTGGCATGGGAGAAAAAAGATATTTT 2864
Db 2647 CTGAGACATTAACAGATAGTTACTTTGGCTTTTGGCATGGGAGAAAAAAGATATTTT 2706
Qy 2865 CAAGATCTTCTCGACGAAGAGAGTCAATGAAGACAAATTTGGCATACTTCACTGATAGC 2924
Db 2707 CTAGATCTTCTCGACGAAGAGAGTCAATGAAGACAAATTTGGCATACTTCACTGATAGC 2766

QY 2925 AAAAATCTGGAGGCACTAAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2984
Db |||||
QY 2967 AAGAATCTGGAGGCACTAAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2826
Db |||||
QY 2985 ATTCTAAATAGCAAGTTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAATG 3044
Db |||||
QY 2827 ATTCTAAATAGCAAGTTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAATG 2886
Db |||||
QY 3045 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 3104
Db |||||
QY 2887 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 2946
Db |||||
QY 3105 TCATTTCAAAAGTGGCCATTCCTGAGGATATGAGATGAGTTGCTTCTCTTTATTTAT 3164
Db |||||
QY 2947 TCATTTCAAAAGTGGCCATTCCTGAGGATATGAGATGAGTTGCTTCTCTTTATTTAT 3006
Db |||||
QY 3165 CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTGAATGAGTTGATACAGAT 3224
Db |||||
QY 3007 CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTGAATGAGTTGATACAGAT 3066
Db |||||
QY 3225 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACACTGGCTACCGAAATTCAGAACTG 3284
Db |||||
QY 3067 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACACTGGCTACCGAAATTCAGAACTG 3126
Db |||||
QY 3285 CCGTTAAGTTTGAGGATTTGACAGGTCTGGAACACATGCTTAATAATTGCTCAAAAATG 3344
Db |||||
QY 3127 CCGTTAAGTTTGAGGATTTGACAGGTCTGGAACACATGCTTAATAATTGCTCAAAAATG 3186
Db |||||
QY 3345 CTTCCTGCTGATATCAGCGAGTAAATAATATCCAGCAACTCAGGAATCCTACTATGAT 3404
Db |||||
QY 3187 CTTCCTGCTGATATCAGCGAGTAAATAATATCCAGCAACTCAGGAATCCTACTATGAT 3246
Db |||||
QY 3405 CCCAACCTGCCACCGGTCTCTAAAGTCTAGTAAACAACTGTAACACAGTAACTGACAAA 3464
Db |||||
QY 3247 CCCAACCTGCCACCGGTCTCTAAAGTCTAGTAAACAACTGTAACACAGTAACTGACAAA 3306
Db |||||
QY 3465 ATCCACAAAGCATATAAGGACAAACAAATATAGTTTCAATCATGCGAGAGAGAA 3524
Db |||||
QY 3307 ATCCACAAAGCATATAAGGACAAACAAATATAGTTTCAATCATGCGAGAGAGAA 3366
Db |||||
QY 3525 ATGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3584
Db |||||
QY 3367 ATGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3426
Db |||||
QY 3585 AGAAAAACCTTAGGAGTTGTTGCTGTAATGACACAACTGCAACATCAATCAATGAT 3644
Db |||||
QY 3427 AGAAAAACCTTAGGAGTTGTTGCTGTAATGACACAACTGCAACATCAATCAATGAT 3486
Db |||||
QY 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCACTCTCC 3704
Db |||||
QY 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCACTCTCC 3546
Db |||||
QY 3705 CAATTTGAACCTGCAAGAGATTCGAAACCGTTTCTCATATGATGATGCTGCAAGAA 3764
Db |||||
QY 3547 CAATTTGAACCTGCAAGAGATTCGAAACCGTTTCTCATATGATGATGCTGCAAGAA 3606
Db |||||
QY 3765 TGGAGGCTTATCAGACAAATGAAGTTTGGACCCATGCTGCTAGCAACATGAT 3824
Db |||||
QY 3607 TGGAGGCTTATCAGACAAATGAAGTTTGGACCCATGCTGCTAGCAACATGAT 3666
Db |||||
QY 3825 ATGTTTACTATATCTCATTTTGTGACAGTAAATGCACTTAAGCGGAGATATTT 3884
Db |||||
QY 3667 ATGTTTACTATATCTCATTTTGTGACAGTAAATGCACTTAAGCGGAGATATTT 3726
Db |||||
QY 3885 CCCAGAGGAGGATACACAAGAGCTAGTCCCAATCGATCAGCTATAGAGATC 3941
Db |||||
QY 3727 CCCAGAGGAGGATACACAAGAGCTAGTCCCAATCGATCAGCTATAGAGATC 3783
Db |||||

RESULT 12

US-09-895-072-21

; Sequence 21, Application US/09895072

; Patent No. US20020025550A1

GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE
; FILE REFERENCE: 210119US00UNT
; CURRENT APPLICATION NUMBER: US/09/895,072
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-895-072-21

Query Match 62.3%; Score 3485.4; DB 9; Length 3621;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAGTCC 356
Db |||||
QY 115 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAGTCC 174
Db |||||
QY 357 TTTCAGAAATCGGCTTTGTCTGCCCATGCGATTGACGTTGTTTACACCTGGGTGAATGCC 416
Db |||||
QY 175 TTTCAGAAATCGGCTTTGTCTGCCCATGCGATTGACGTTGTTTACACCTGGGTGAATGCC 234
Db |||||
QY 417 ACAGATCTTGAATCTAGGAACTACAGCAGGTACAGAACTAGAGAGAGAGAGAG 476
Db |||||
QY 235 ACAGATCTTGAATCTAGGAACTACAGCAGGTACAGAACTAGAGAGAGAGAGAG 294
Db |||||
QY 477 AAAGCAATGAGAGAAATCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 536
Db |||||
QY 295 AAAGCAATGAGAGAAATCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 354
Db |||||
QY 537 CAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGACTGACCCAGCC 596
Db |||||
QY 355 CAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGACTGACCCAGCC 414
Db |||||
QY 597 CTGCCAGCCACATCACCCTGAGAGACGTCGATCTCTTTATCTCTCTCTCTCTCTCT 656
Db |||||
QY 415 CTGCCAGCCACATCACCCTGAGAGACGTCGATCTCTTTATCTCTCTCTCTCTCTCT 474
Db |||||
QY 657 AGTGACATTTCAATGTTGCAAAACCAAAAAACCCCTTACCAATGCTCAGTTGTTGTT 716
Db |||||
QY 475 AGTGACATTTCAATGTTGCAAAACCAAAAAACCCCTTACCAATGCTCAGTTGTTGTT 534
Db |||||
QY 717 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 776
Db |||||
QY 535 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 594
Db |||||
QY 777 CAGACAGTATGGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGATTAAGTCTAATG 836
Db |||||
QY 595 CAGACAGTATGGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGATTAAGTCTAATG 654
Db |||||
QY 837 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAA 896
Db |||||
QY 655 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAA 714
Db |||||
QY 897 ACAAAATTCGAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTCTGATTCAGAGGCC 956
Db |||||
QY 715 ACAAAATTCGAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTCTGATTCAGAGGCC 774
Db |||||
QY 957 AGTGTAGCGCTTTCAAAACCTGAATTAACCCCAAGGATTTTCAAGAAATTAAGTCAACT 1016
Db |||||
QY 775 AGTGTAGCGCTTTCAAAACCTGAATTAACCCCAAGGATTTTCAAGAAATTAAGTCAACT 834
Db |||||
QY 1017 AAGAGAACATGACCATGATGGAAGAACTGACCACTAGCTCTGATATTTATTATGG 1076
Db |||||
QY 835 AAGAGAACATGACCATGATGGAAGAACTGACCACTAGCTCTGATATTTATTATGG 894
Db |||||

QY	1077	GATCTGAGGCCATCAGCCAGCTTAAGCAGAGATGAAGACATCTCTGCCAGTCGTTTGA	1136
Db	895	GATCTGAGGCCATCAGCCAGCTTAAGCAGAGATGAAGACATCTCTGCCAGTCGTTTGA	954
QY	1137	GATAACGAAGAACTCAGGTACTCATTTGGATCTATCGAGAGGCATGCACCATGGTTCGG	1196
Db	955	GATACGGAAGAACTCAGGTACTCATTTGGATCTATCGAGAGGCATGCACCATGGTTCGG	1014
QY	1197	AATATTTTCAATTTGTCACCAACGGGAGATTCATCTCTGGCTGAACCTTGAACATCTCGA	1256
Db	1015	AATATTTTCAATTTGTCACCAACGGGAGATTCATCTCTGGCTGAACCTTGAACATCTCGA	1074
QY	1257	GTGCAATAAGTAACACACAGGATGTTTTTCGAAATTTGAGCCACATTCGCCTACTTAGT	1316
Db	1075	GTGCAATAAGTAACACACAGGATGTTTTTCGAAATTTGAGCCACATTCGCCTACTTAGT	1134
QY	1317	TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCGAGAAAGTTTATTAC	1376
Db	1135	TCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCGAGAAAGTTTATTAC	1194
QY	1377	CTAATGATGATGTCATGTTTTGGGAAGATCTCTGGCCAGATGATTTTTACAGTCACTCC	1436
Db	1195	CTAATGATGATGTCATGTTTTGGGAAGATGTCCTGGCCAGATGATTTTTACAGTCACTCC	1254
QY	1437	AAAGGCCAGAAGGTTTTATTGACATGGCTGTGCCAAATGTGTGCCGAGGGCTGCCAGGT	1496
Db	1255	AAAGGCCAGAAGGTTTTATTGACATGGCTGTGCCAAATGTGTGCCGAGGGCTGCCAGGT	1314
QY	1497	TCTCGATTAAGATGGCTATTGTGACAAGGCTCTAATAATTCAGCCTCGGATGGGAT	1556
Db	1315	TCTCGATTAAGATGGCTATTGTGACAAGGCTCTAATAATTCAGCCTCGGATGGGAT	1374
QY	1557	GGTGGGATTTGCTCTGGAAACAGTGGAGGGAGTGGCTATATTGACGAGGTGGAGTACT	1616
Db	1375	GGTGGGATTTGCTCTGGAAACAGTGGAGGGAGTGGCTATATTGACGAGGTGGAGTACT	1434
QY	1617	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAGTGTCTCTAC	1676
Db	1435	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAGTGTCTCTAC	1494
QY	1677	TGTAATCAGGGATGTCGGAATTCCTGGCTCGCTGATAGTATCTGTGACCAAGCATGCAAT	1736
Db	1495	TGTAATCAGGGATGTCGGAATTCCTGGCTCGCTGATAGTATCTGTGACCAAGCATGCAAT	1554
QY	1737	GTCTTGCTGTGGGTTTGATGCTGGGAGCTGTGGGGAAGATCAATTTTCATGAATGTAT	1796
Db	1555	GTCTTGCTGTGGGTTTGATGCTGGGAGCTGTGGGGAAGATCAATTTTCATGAATGTAT	1614
QY	1797	AAAGTGATCTCTCCCAACACAGACTCACTATATATTTCAAAAGGTGAATGCTGCCT	1856
Db	1615	AAAGTGATCTCTCTCCCAACACAGACTCACTATATATTTCAAAAGGTGAATGCTGCCT	1674
QY	1857	TATTTTCAGCTTTGCAAGTAGCCAAAAGAGAGGTGAAGTGGCTATAGTGAACAATCCA	1916
Db	1675	TATTTTCAGCTTTGCAAGTAGCCAAAAGAGAGGTGAAGTGGCTATAGTGAACAATCCA	1734
QY	1917	ATAATTCGATGCTTTCTATTCGCACAGTGGAAAACCATCCACTCATATATCCACAGT	1976
Db	1735	ATAATTCGATGCTTTCTATTCGCACAGTGGAAAACCATCCACTCATATATCCACAGT	1794
QY	1977	GGAAATGAATGCCACCAACAATATTTTAAATCTCAAGTTTCAAAAATACAAACGATGAAG	2036
Db	1795	GGAAATGAATGCCACCAACAATATTTTAAATCTCAAGTTTCAAAAATACAAACGATGAAG	1854
QY	2037	TTCAAAATCGAGATAACAGTGGAGGTGGACACAGGGGGGACCAAACTGAATCTACG	2096
Db	1855	TTCAAAATCGAGATAACAGTGGAGGTGGACACAGGGGGGACCAAACTGAATCTACG	1914
QY	2097	GGCCAGAGGGTTTACGAAAATTTAGTTAGTCCCATAAACATCTTCTCCAGAGGGCGAAATC	2156
Db	1915	GGCCAGAGGGTTTACGAAAATTTAGTTAGTCCCATAAACATCTTCTCCAGAGGGCGAAATC	1974
QY	2157	CTTTTTGAGGATATTTCCCAAGAAAAACGCTTCCGAAAGTTTAAAGACATGATGTTAAAC	2216
Db	1975	CTTTTTGAGGATATTTCCCAAGAAAAACGCTTCCGAAAGTTTAAAGACATGATGTTAAAC	2034
QY	2217	TCACAGAGGAGCCCGAGAGAGGTGAATAATTTCCCTGGTAAATATTTCACTCCTTCCA	2276
Db	2035	TCACAGAGGAGCCCGAGAGAGGTGAATAATTTCCCTGGTAAATATTTCACTCCTTCCA	2094
QY	2277	AAAGACGCCCCAGTTGAGTCTCAATATCTTTGGATTTTGCACCTTGGAAATATTTCACT	2336
Db	2095	AAAGACGCCCCAGTTGAGTCTCAATATCTTTGGATTTTGCACCTTGGAAATATTTCACT	2154
QY	2337	TTGAAGGATACAAATTTTGCACAGTCAGCCTTGTCTGAGATCAATTTCTTGATCACTCACAG	2396
Db	2155	TTGAAGGATACAAATTTTGCACAGTCAGCCTTGTCTGAGATCAATTTCTTGATCACTCACAG	2214
QY	2397	CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAAAATAATGACAGTTTGGTGGCT	2456
Db	2215	CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAAAATAATGACAGTTTGGTGGCT	2274
QY	2457	CCACAGAAAAAACAGGTTTCATAAAGCATCTTGCACACACAGCTTGAAGAGTCTCGAAGA	2516
Db	2275	CCACAGAAAAAACAGGTTTCATAAAGCATCTTGCACACACAGCTTGAAGAGTCTCGAAGA	2334
QY	2517	TTGCAGAGGTTGACTTTTCTGCAGTGAGTGTAAGAAGTGAATGGTTCATGACACAGGGTCAG	2576
Db	2335	TTGCAGAGGTTGACTTTTCTGCAGTGAGTGTAAGAAGTGAATGGTTCATGACACAGGGTCAG	2394
QY	2577	AATCCACCCCTGCAGTTGGAGACCCACAGCAAGATTTAGAGTGGAAATCTCACACCAAAAA	2636
Db	2395	AATCCACCCCTGCAGTTGGAGACCCACAGCAAGATTTAGAGTGGAAATCTCACACCAAAAA	2454
QY	2637	ACCATAGCCGGAATGTGCAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC	2696
Db	2455	ACCATAGCCGGAATGTGCAAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGC	2514
QY	2697	CAGATGACAAAAAGAAAGAAATCACAAGGGAAGAAAGAAACAGTAGAATGGAGGAA	2756
Db	2515	CAGATGACAAAAAGAAAGAAATCACAAGGGAAGAAAGAAACAGTAGAATGGAGGAA	2574
QY	2757	AATCCTGAAAAATCACATAGCGTTACTGGAAGTGTACTTGGGAAGAAAGCTGCAGCATTAC	2816
Db	2575	AATCCTGAAAAATCACATAGCGGTACTGGAAGTGTACTTGGGAAGAAAGCTGCAGCATTAC	2634
QY	2817	ACAGATAGTTACTTTGGGCTTTTTCGATGGGAAAAAAGTATTTTCCAGATCTTCTC	2876
Db	2635	ACAGATAGTTACTTTGGGCTTTTTCGATGGGAAAAAAGTATTTTCTAGATCTTCTC	2694
QY	2877	GACGAAGAGAGTCATTTGAAGACACAAATTTGGCATTTTCCATTTGGGAAAAAAGTATTT	2936
Db	2695	GACGAAGAGAGTCATTTGAAGACACAAATTTGGCATTTTCCATTTGGGAAAAAAGTATTT	2754
QY	2937	AGGCAACTTAAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGC	2996
Db	2755	AGGCAACTTAAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGC	2814
QY	2		

1497 TCCTGGATTAAAGGATGGCTATTGTGACAAAGGCTGTGTAATAATTCAGCCTCGGATGGGAT 1556
1315 TCCTGGATTAAAGGATGGCTATTGTGACAAAGGCTGTGTAATAATTCAGCCTCGGATGGGAT 1374
1557 GGTGGGATTCCTCTGGAAACAGTGGAGGGAGTGGCTATATTCAGAGAGGAGTGGAGTACT 1616
1375 GGTGGGATTCCTCTGGAAACAGTGGAGGGAGTGGCTATATTCAGAGAGGAGTGGAGTACT 1434
1617 GGGAGTATTCAGGATTCGACACCCCTGGCAGCTTTGGTGGAGGAATAAACAAGTGTCTCTTAC 1676
1435 GGGAGTATTCAGGATTCGACACCCCTGGCAGCTTTGGTGGAGGAATAAACAAGTGTCTCTTAC 1494
1677 TGTAAATCAGGATTCGGAATTCCTGGCTCGCTGATGAAGTTCGTGACCAAGCATGCAAT 1736
1495 TGTAAATCAGGATTCGGAATTCCTGGCTCGCTGATGAAGTTCGTGACCAAGCATGCAAT 1554
1737 GTCTGTCTGTGGGTTTGTGCTGGCGACTGTGGCAAGATCATTTTCATGAATGTAT 1796
1555 GTCTGTCTGTGGGTTTGTGCTGGCGACTGTGGCAAGATCATTTTCATGAATGTAT 1614
1797 AAAGTGTATCTCTCCCAAAACCAAGCTCACTATATTTATTCAAAAGGTGAATGCCGTCT 1856
1615 AAAGTGTATCTCTCCCAAAACCAAGCTCACTATATTTATTCAAAAGGTGAATGCCGTCT 1674
1857 TATTTAGCTTTGCAAGATAGCCAAAGAGGATTTGAAGTGCCTATAGTGACATCCA 1916
1675 TATTTAGCTTTGCAAGATAGCCAAAGAGGATTTGAAGTGCCTATAGTGACATCCA 1734
1917 ATAAATCGACATGCTTCTATTTGCCAACCAAGTGAAGAACCAATCCACCTCAATATGCACAGT 1794
1735 ATAAATCGACATGCTTCTATTTGCCAACCAAGTGAAGAACCAATCCACCTCAATATGCACAGT 1794
1977 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATCAAAACGATGAGAG 2036
1795 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATCAAAACGATGAGAG 1854
2037 TTCAAAATGACAGATAACAGTGGAGGTGGACACAAAGGAGGAGCAAAACCTGAATTTCTAG 2096
1855 TTCAAAATGACAGATAACAGTGGAGGTGGACACAAAGGAGGAGCAAAACCTGAATTTCTAG 1914
2097 GCCAGAGGAGGTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 2156
1915 GCCAGAGGAGGTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 1974
2157 CTTTTCGAGGATATCCCAAGAAACAGCTTCCCGAAGTTTAAAGAGACATGATTTAAC 2216
1975 CTTTTCGAGGATATCCCAAGAAACAGCTTCCCGAAGTTTAAAGAGACATGATTTAAC 2034
2217 TCACAAAGGAGAGCCCAAGGAGGTGAAATTCCTCGTGAATAATTTCACTCTCCCA 2276
2035 TCACAAAGGAGAGCCCAAGGAGGTGAAATTCCTCGTGAATAATTTCACTCTCCCA 2094
2277 AAAGAGCCCGATTTGAGTCTCAATACCTTGGATTTGCAACTGGAATCAATGATTTAAC 2336
2095 AAAGAGCCCGATTTGAGTCTCAATACCTTGGATTTGCAACTGGAATCAATGATTTAAC 2154
2337 TTGAAGGATACAAATTTGTCCAGCTCAGCTTGTGAGATCATTTCTGTGAATCACTCACAG 2396
2155 TTGAAGGATACAAATTTGTCCAGCTCAGCTTGTGAGATCATTTCTGTGAATCACTCACAG 2214
2397 CATGCTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTTGGTGGCT 2456
2215 CATGCTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTTGGTGGCT 2274
2457 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGAGGTCTCGAAGA 2516
2275 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGAGGTCTCGAAGA 2334
2517 TTGCAGAGGTTGACTTTTCTGCAAGTGAAGTGAATGGTCAATGACAGGCTCAG 2576
2335 TTGCAGAGGTTGACTTTTCTGCAAGTGAAGTGAATGGTCAATGACAGGCTCAG 2394
2577 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTTAGAGTGGAAATTCACACCCCAAAA 2636

2395 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTTAGAGTGGAAATTCACACCCCAAAA 2454
2637 ACCATAGCGGAAATGTGACAAAGAAAGCCCCATCTCTGATTGTTCACCTGGAAAGC 2696
2455 ACCATAGCGGAAATGTGACAAAGAAAGCCCCATCTCTGATTGTTCACCTGGAAAGC 2514
2697 CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAAAGACAGTAGAATGGAGAA 2756
2515 CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAAAGACAGTAGAATGGAGAA 2574
2757 AATCTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCAATTAC 2816
2575 AATCTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCAATTAC 2634
2817 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTCCAAAGATCTTCTC 2876
2635 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTCCTAGATCTTCTC 2694
2877 GACGAAGAGAGTCAATTTGAAGACACAATTTGSCATACCTCTCATGATAGCAAAATACCTGGG 2936
2695 GACGAAGAGAGTCAATTTGAAGACACAATTTGSCATACCTCTCATGATAGCAAAATACCTGGG 2754
2937 AGGCAACTTAAAGATACATTTGCAAGATTCCTCTCAGATATGTAATAAAATTTCTAAATPAGC 2996
2755 AGGCAACTTAAAGATACATTTGCAAGATTCCTCTCAGATATGTAATAAAATTTCTAAATPAGC 2814
2997 AAGTTTGGATTCACATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 3056
2815 AAGTTTGGATTCACATCGCGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 2874
3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGAGCTCATTTTCAAAA 3116
2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGAGCTCATTTTCAAAA 2934
3117 GTGCGCCATTCGAGGATATGAGTGTGCTTCTCTTATTTATTTATCTCATGAGTGCA 3176
2935 GTGCGCCATTCGAGGATATGAGTGTGCTTCTCTTATTTATTTATCTCATGAGTGCA 2994
3177 GTGCGCCATTCGAGGATATGAGTGTGCTTCTCTTATTTATTTATCTCATGAGTGCT 3236
2995 GTGCGCCATTCGAGGATATGAGTGTGCTTCTCTTATTTATTTATCTCATGAGTGCT 3054
3237 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCAGAACTGCGGTTAAGTTTG 3296
3055 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCAGAACTGCGGTTAAGTTTG 3114
3297 CAGGATTTGACAGGCTTGGAGACATGCTTAATTAATTTGCTCAAAATGCTTCTGCTGAT 3356
3115 CAGGATTTGACAGGCTTGGAGACATGCTTAATTAATTTGCTCAAAATGCTTCTGCTGAT 3174
3357 ATCAGCAGCTTAAATATATTTCCCAACCTCAGAAATCTCTATGATCCCAACCTGCCA 3416
3175 ATCAGCAGCTTAAATATATTTCCCAACCTCAGAAATCTCTATGATCCCAACCTGCCA 3234
3417 CCGGTCACTAAAGTCTAGTAAACAACTGTAACCCAGTAACTGACAAATCCCAAGAGCA 3476
3235 CCGGTCACTAAAGTCTAGTAAACAACTGTAACCCAGTAACTGACAAATCCCAAGAGCA 3294
3477 TATTAAGGACAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAA 3536
3295 TATTAAGGACAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAA 3354
3537 ATGATTCCTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGATGACAAAGAACCT 3596
3355 ATGATTCCTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGATGACAAAGAACCT 3414
3597 AGGAGTTTGTTCCTGAATGACAAATGACCAATGACCAATGACCAATGACCAAGCT 3656
3415 AGGAGTTTGTTCCTGAATGACAAATGACCAATGACCAATGACCAAGCT 3474
3657 AAGGCTGTCTCAGGACTTCTATGAATCCATGTTCCTCCATACCTTCCCAATTTGAACTG 3716

Db	3475	AAGCTGTTCTCAGGGACCTTCATGAAATCCATGTTCCTCCCATACCTTCCCAATTGTGAAC	3534
Qy	3717	CCAAAGAGAGATCGAAACCGTTTCCTTCATATGATGAGCTGCAGGAATCGAGGGCTTAT	3776
Db	3535	CCAAGAGAGATCGAAACCGTTTCCTTCATATGATGAGCTGCAGGAATCGAGGGCTTAT	3594
Qy	3777	CGACACAAATTGAAGTTTGTGAC	3799
Db	3595	CGACACAAATTGAAGTAGTAGTC	3617

RESULT 14

```

US-10-306-686-21
; Sequence 21, Application US/10306596
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCNACase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-306-686-21

```

```
Query Match      62.3%; Score 3485.4; DB 15; Length 3621;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

QY	297	AGCGGAGATCAATACCAATGTTTGTGTTGATTCCTATAGAGACAATATGCTGGAAAGTCC	356
DB	115	AGCGGAGATCAATACCAATGTTTGTGTTGATTCCTATAGAGACAATATGCTGGAAAGTCC	174
QY	357	TTTCAGAATCGGCTTTTCTGCGCCCATGCCGATTGACGTTGTTTACACCTCGGGTGAATGGC	416
DB	175	TTTCAGAATCGGCTTTTCTGCGCCCATGCCGATTGACGTTGTTTACACCTCGGGTGAATGGC	234
QY	417	ACAGATCTTGAACCTACTGTAAGGAACCTACAGCAGGTACAGAGAACAGATGGAGGAGGACGAG	476
DB	235	ACAGATCTTGAACCTACTGTAAGGAACCTACAGCAGGTACAGAGAACAGATGGAGGAGGACGAG	294
QY	477	AAAGCAATAGAGAGAAATCCCTTGGGAAAAACACAACCGGAACTCTACTAAGAGAGAGTGAGAAG	536
DB	295	AAAGCAATAGAGAGAAATCCCTTGGGAAAAACACAACCGGAACTCTACTAAGAGAGAGTGAGAAG	354
QY	537	CAGTTAGAGTGTGTTGCTAAACACACTGCAATTAAGTGGCAATGCTTGTTACTGGACCCAGCC	596
DB	355	CAGTTAGAGTGTGTTGCTAAACACACTGCAATTAAGTGGCAATGCTTGTTACTGGACCCAGCC	414
QY	597	CTGCCAGCCAAATCAGCCCTGAAGGACGTGCGCATCTCTTTATCCCTCTTTTTCATCTGCCC	656
DB	415	CTGCCAGCCAAATCAGCCCTGAAGGACGTGCGCATCTCTTTATCCCTCTTTTTCATCTGCCC	474
QY	657	AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCTCAGTTGTTGTTT	716
DB	475	AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCTCAGTTGTTGTTT	534
QY	717	TTTGACAGTACTAAGGATGTTGAAGATGCCCCACCTCTGCACTGCTTTTAAAGGAAATAGCAGA	776
DB	535	TTTGACAGTACTAAGGATGTTGAAGATGCCCCACCTCTGCACTGCTTTTAAAGGAAATAGCAGA	594
QY	777	CAGACAGTATGGAGGGGTACTTGCACCAACAGATAAAGAGTCCCTGGATTAGTGCTAATG	836
DB	595	CAGACAGTATGGAGGGGTACTTGCACCAACAGATAAAGAGTCCCTGGATTAGTGCTAATG	654

Db 1735 ATAAATCGCATGCTTCTATTGGCCAAAGTGGAAAACCATCCACCTCATATATGSCACAGT 1794
Qy 1977 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTTCAAAATACAAACGATGAAGAG 2036
Db 1795 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAAGAG 1854
Qy 2037 TTCAAAATGACAGTAAACAGTGGAGGTGGACACAAGGGAGGGACCAAACTGAATTTACG 2096
Db 1855 TTCAAAATGACAGTAAACAGTGGAGGTGGACACAAGGGAGGGACCAAACTGAATTTACG 1914
Qy 2097 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATACACTCTTCCAGAGGGCGAAATC 2156
Db 1915 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATACACTCTTCCAGAGGGCGAAATC 1974
Qy 2157 CTTTTTGGAGGATATCCCAAGAAAACCGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2216
Db 1975 CTTTTTGGAGGATATCCCAAGAAAACCGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2034
Qy 2217 TCAACAGAGGAGCCACAGGAGAGGTGAAAATTTCCCTGTGTAATATTTCACTCTCTCCA 2276
Db 2035 TCAACAGAGGAGCCACAGGAGAGGTGAAAATTTCCCTGTGTAATATTTCACTCTCTCCA 2094
Qy 2277 AAAGACCCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGACATCACT 2336
Db 2095 AAAGACCCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGACATCACT 2154
Qy 2337 TTGAAGGATACAAATTTGTCAGTACGCTTCTGAGATCATTTCTGATGAATCACTCACAG 2396
Db 2155 TTGAAGGATACAAATTTGTCAGTACGCTTCTGAGATCATTTCTGATGAATCACTCACAG 2214
Qy 2397 CATCTAAATATAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT 2456
Db 2215 CATCTAAATATAAATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT 2274
Qy 2457 CCACAGAAAACAGGTTTCATAAAGCATCTTCCAAACAGCTTAGGAGTGTCTGAAGA 2516
Db 2275 CCACAGAAAACAGGTTTCATAAAGCATCTTCCAAACAGCTTAGGAGTGTCTGAAGA 2334
Qy 2517 TTGCAGAGGTGACTTTTCTCGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2576
Db 2335 TTGCAGAGGTGACTTTTCTCGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2394
Qy 2577 AATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGGAATCTCACACCCCAAAA 2636
Db 2395 AATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGGAATCTCACACCCCAAAA 2454
Qy 2637 ACCATAGGCGGAATGTGACAAAAGAAAGCCCCCATCTCTGATTTCTCAGTGGAAAGC 2696
Db 2455 ACCATAGGCGGAATGTGACAAAAGAAAGCCCCCATCTCTGATTTCTCAGTGGAAAGC 2514
Qy 2697 CAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAGAACAGTAGATGGAAGAA 2756
Db 2515 CAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAAAGAGAACAGTAGATGGAAGAA 2574
Qy 2757 AATGCTGAAAATCACATAGCGGTTACTGAAGTGTACTTGGAAAGAAAGCTGACAGCAATAC 2816
Db 2575 AATGCTGAAAATCACATAGCGGTTACTGAAGTGTACTTGGAAAGAAAGCTGACAGCAATAC 2634
Qy 2817 ACAGATAGTTACTTGGGCTTTTGGCAATGGGAGAAAAGATTTTCCAGAGATCTTCTC 2876
Db 2635 ACAGATAGTTACTTGGGCTTTTGGCAATGGGAGAAAAGATTTTCTGATCTTCTC 2694
Qy 2877 GACGAGAGAGAGTCAATGAAGACACAAATTTGGCACTTCTGATGAGCAAAAATCTGGG 2936
Db 2695 GACGAGAGAGTCAATGAAGACACAAATTTGGCACTTCTGATGAGCAAAAATCTGGG 2754
Qy 2937 AGGCACTAAAGATACATTTGAGATTCCTCAGATATGTAATATAAATTTCTAATAGC 2996
Db 2755 AGGCACTAAAGATACATTTGAGATTCCTCAGATATGTAATATAAATTTCTAATAGC 2814
Qy 2997 AAGTTTGGATTCATCGCGGAAAGTCCCTGCTCACATGCTCCATGATTTGACCGGATT 3056

Db 2815 AAGTTTGGATTCATCGCGGAAAGTCCCTGCTCACATGCTCCATGATTTGACCGGATT 2874
Qy 3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGAGCTCATTTCAAAA 3116
Db 2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAAGAGCTCATTTCAAAA 2934
Qy 3117 GTGCGCCATTTCTGAGGATATGCAAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCA 3176
Db 2935 GTGCGCCATTTCTGAGGATATGCAAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCA 2994
Qy 3177 GTGCGCCATTTCTGAGGATATGCAAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCA 3236
Db 2995 GTGCGCCATTTCTGAGGATATGCAAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCA 3054
Qy 3237 TTGCTGACAGAGAAATCCGAAACCTGGCTACCAAGATTTCAAGATTCAGATCAATCTGGTGTG 3296
Db 3055 TTGCTGACAGAGAAATCCGAAACCTGGCTACCAAGATTTCAAGATTCAGATCAATCTGGTGTG 3114
Qy 3297 CAGGATTTGACAGTCTGGACACATGCTAATAATTTGCTCAAAAATGCTTCTGCTGAT 3356
Db 3115 CAGGATTTGACAGTCTGGACACATGCTAATAATTTGCTCAAAAATGCTTCTGCTGAT 3174
Qy 3357 ATCAGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTACTATGATCCCACTGCCA 3416
Db 3175 ATCAGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTACTATGATCCCACTGCCA 3234
Qy 3417 CCGTCTCAATAAGTCTAGTAAACAACTGTAAACCACTGTAACCAAAATGCTCAAAAATGCTCAAAA 3476
Db 3235 CCGTCTCAATAAGTCTAGTAAACAACTGTAAACCACTGTAACCAAAATGCTCAAAAATGCTCAAAA 3294
Qy 3477 TATAAGACAAAACAAATATAGCTTTGAAATCATGGAAGAAAGAAATGCTTTTAAA 3536
Db 3295 TATAAGACAAAACAAATATAGCTTTGAAATCATGGAAGAAAGAAATGCTTTTAAA 3354
Qy 3537 ATGATTCGTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCT 3596
Db 3355 ATGATTCGTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCT 3414
Qy 3597 AGGAAGTTTGTTCCTGAAATGACAACTTGACCACAAATCATAAAGATGCTCAGACAGTG 3656
Db 3415 AGGAAGTTTGTTCCTGAAATGACAACTTGACCACAAATCATAAAGATGCTCAGACAGTG 3474
Qy 3657 AAGCTGTTCCTCAGGACTTCTATGAATCCATGTTCCTCCCATACCTTCCCAATTTGAACTG 3716
Db 3475 AAGCTGTTCCTCAGGACTTCTATGAATCCATGTTCCTCCCATACCTTCCCAATTTGAACTG 3534
Qy 3717 CCAAGAGAGTATCGAAACCGTTTCTTCATATGATGAGCTGAGGAAATGAGGGCTTAT 3776
Db 3535 CCAAGAGAGTATCGAAACCGTTTCTTCATATGATGAGCTGAGGAAATGAGGGCTTAT 3776
Qy 3777 CGACAAAATTTGAAAGTTTGGAC 3799
Db 3595 CGACAAAATTTGAAAGTAGTAGTC 3617

RESULT 15
US-10-023-888-1
; Sequence 1, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 2035150577
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: hybrid
US-10-023-888-1

Query Match					61.9%; Score 3465.6; DB 15; Length 3600;				
Best Local Similarity					99.5%; Pred. No. 0;				
Matches 3477; Conservative					0; Mismatches 19; Indels 0; Gaps 0;				
Qy	297	AGCGAGATCAATACCACTGTTTGTGTTGATTCCTATAGAGACAATAATGCTGGAAGTCC	356						
Db	103	AGCGAGATCAATACCACTGTTTGTGTTGATTCCTATAGAGACAATAATGCTGGAAGTCC	162						
Qy	357	TTTCAGAAATCGGCTTTGTCTCCCATGCGCGATTGACGTTGTTTACACTGGTGAATGGC	416						
Db	163	TTTCAGAAATCGGCTTTGTCTCCCATGCGCGATTGACGTTGTTTACACTGGTGAATGGC	222						
Qy	417	ACAGATCTTGAACTACTGAGGAACACTACAGCAGTTCAGAGAAACAGATGGAGGAGCAG	476						
Db	223	ACAGATCTTGAACTACTGAGGAACACTACAGCAGTTCAGAGAAACAGATGGAGGAGCAG	282						
Qy	477	AAAGCAATAGAGAAATCCCTTGGGAAAACAAACGGAACCTACTAAGAAAGAGTGAGAAG	536						
Db	283	AAAGCAATAGAGAAATCCCTTGGGAAAACAAACGGAACCTACTAAGAAAGAGTGAGAAG	342						
Qy	537	CAGTTAGAGTCTTTGCTTAACACATGCAATTAAGTGGCCATGCTTGACTGGACCCAGCC	596						
Db	343	CAGTTAGAGTCTTTGCTTAACACATGCAATTAAGTGGCCATGCTTGACTGGACCCAGCC	402						
Qy	597	CTGCCAGCAACATCACCCCTGAAGGACGTGCCATCTCTTTATCCTCTCTTTTCATTTCTGCC	656						
Db	403	CTGCCAGCAACATCACCCCTGAAGGACGTGCCATCTCTTTATCCTCTCTTTTCATTTCTGCC	462						
Qy	657	AGTGACATTTTCAATGTTGGAACACCAAAAACCCCTTACCAATGCTCAGTTGTTGTT	716						
Db	463	AGTGACATTTTCAATGTTGGAACACCAAAAACCCCTTACCAATGCTCAGTTGTTGTT	522						
Qy	717	TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGA	776						
Db	523	TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGA	582						
Qy	777	CAGACAGTATCGAGGGGTACTTGACAAACAGATTAAGAGTCCCTGGATAGTGTCTAATG	836						
Db	583	CAGACAGTATCGAGGGGTACTTGACAAACAGATTAAGAGTCCCTGGATAGTGTCTAATG	642						
Qy	837	CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA	896						
Db	643	CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA	702						
Qy	897	ACAAAATGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTGTATTCAGAGGCC	956						
Db	703	ACAAAATGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTGTATTCAGAGGCC	762						
Qy	957	AGTGAGCGCTTTAAACTGAAATAACCCGAGGATTTCAAGAAATTAAGTGAAGCAACT	1016						
Db	763	AGTGAGCGCTTTCTAAACTGAAATAACCCGAGGATTTCAAGAAATTAAGTGAAGCAACT	822						
Qy	1017	AAGAAGACATGACCATTTGATGGAAGAAAGAACTGACCAATAGTCTCGCATATTTATATGG	1076						
Db	823	AAGAAGACATGACCATTTGATGGAAGAAAGAACTGACCAATAGTCTCGCATATTTATATGG	882						
Qy	1077	GATCTGAGCGCATCAGCCATCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTTGAA	1136						
Db	883	GATCTGAGCGCATCAGCCATCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTTGAA	942						
Qy	1137	GATACGAGAACTGAGTACTCATGCGATCTATCGAGGGCATGCACATGCGGTTTCGG	1196						
Db	943	GATACGAGAACTGAGTACTCATGCGATCTATCGAGGGCATGCACATGCGGTTTCGG	1002						
Qy	1197	AATATTTTTCATTTGTCACCAACGGGAGATTCCTCTGGCTGAACCTTTGACAATCCCTCGA	1256						
Db	1003	AATATTTTTCATTTGTCACCAACGGGAGATTCCTCTGGCTGAACCTTTGACAATCCCTCGA	1062						
Qy	1257	GTGACAAATAGTAAACACACAGAGTCTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGT	1316						
Db	1063	GTGACAAATAGTAAACACACAGAGTCTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGT	1122						
Qy	1317	TCACCTGTATTTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCCGAAGTTTATTTAC	1376						

Db	1123	TCACCTGTATTTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTTAC	1182						
Qy	1377	CTAAATGATGATGTATGTTTGGGAAGGATGTCTGCCAGATGATTTTACAGTCACTCC	1436						
Db	1183	CTAAATGATGATGTATGTTTGGGAAGGATGTCTGCCAGATGATTTTACAGTCACTCC	1242						
Qy	1437	AAAGCCAGAAAGTTTATTTGACATGCTGTCGCCAATCTGCGAGGGCTGCCAGGT	1496						
Db	1243	AAAGCCAGAAAGTTTATTTGACATGCTGTCGCCAATCTGCGAGGGCTGCCAGGT	1302						
Qy	1497	TCTCGATTAAGGATGGCTTATTTGACAAAGGCTTCTTAATAATTCAGCCTGCGATTTGGAT	1556						
Db	1303	TCTCGATTAAGGATGGCTTATTTGACAAAGGCTTCTTAATAATTCAGCCTGCGATTTGGAT	1362						
Qy	1557	GGTGGGATGCTCTGGAACACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGTACT	1616						
Db	1363	GGTGGGATGCTCTGGAACACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGTACT	1422						
Qy	1617	GGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1676						
Db	1423	GGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1482						
Qy	1677	TGTAATCAGGATGTGCGAAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGCAAT	1736						
Db	1483	TGTAATCAGGATGTGCGAAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGCAAT	1542						
Qy	1737	GTCTTGTCTGTGGTTTTCATGCTGGCAGTCTGGGCAAGATCATTTTTCATGAATGTAT	1796						
Db	1543	GTCTTGTCTGTGGTTTTCATGCTGGCAGTCTGGGCAAGATCATTTTTCATGAATGTAT	1602						
Qy	1797	AAAGTGATCTCTTCCCAACACAGACTCACTATATTTATTTCCAAAAGGTGAATTCCTGCT	1856						
Db	1603	AAAGTGATCTCTTCCCAACACAGACTCACTATATTTATTTCCAAAAGGTGAATTCCTGCT	1662						
Qy	1857	TATTTTCAGCTTTTCAGAGTACGCAAGAGAGGTTGAGGTGCCTATAGTCACATCCA	1916						
Db	1663	TATTTTCAGCTTTTCAGAGTACGCAAGAGAGGTTGAGGTGCCTATAGTCACATCCA	1722						
Qy	1917	ATAAATTCGACATGCTCTTATTTGCCAACAGTGGAAAAACCATCCACTCATATGACACGT	1976						
Db	1723	ATAAATTCGACATGCTCTTATTTGCCAACAGTGGAAAAACCATCCACTCATATGACACGT	1782						
Qy	1977	GGATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAAGAG	2036						
Db	1783	GGATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAAGAG	1842						
Qy	2037	TTCAAAATCAGATTAACAGTGGAGTGGACACAAAGGGGACCAAACTGAATTTCTAGC	2096						
Db	1843	TTCAAAATCAGATTAACAGTGGAGTGGACACAAAGGGGACCAAACTGAATTTCTAGC	1902						
Qy	2097	GCCAGAGGGTTAGGAAATTTAGTTAGTCCCATTAACACTTTCTTCCAGAGCGGAAATC	2156						
Db	1903	GCCAGAGGGTTAGGAAATTTAGTTAGTCCCATTAACACTTTCTTCCAGAGCGGAAATC	1962						
Qy	2157	CTTTTTCAGGATATTTCCGAAAGAAACCGTTTCCGAGTTTAAAGACATGATTTAC	2216						
Db	1963	CTTTTTCAGGATATTTCCGAAAGAAACCGTTTCCGAGTTTAAAGACATGATTTAC	2022						
Qy	2217	TCAAACAGAGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTCACTCTCTTCA	2276						
Db	2023	TCAAACAGAGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTCACTCTCTTCA	2082						
Qy	2277	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAACATGGAGACATCACT	2336						
Db	2083	AAAGAGCCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAACATGGAGACATCACT	2142						
Qy	2337	TTGAAAGGATACAAATTTGTCGAAGTCAAGCTTGTGAGATCATTTCTGTGAATCAACAG	2396						
Db	2143	TTGAAAGGATACAAATTTGTCGAAGTCAAGCTTGTGAGATCATTTCTGTGAATCAACAG	2202						
Qy	2397	CATGCTAAATAAAAATCAAGCTAATAACAGATGAAGAAACAAATGAACAGTTTGTGCT	2456						

Db 2203 CATGCTAAATATAAATCAAGCTATATTAACAGATGAACAAATGACAGCTTTGGTGGCT 2262
Qy 2457 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2516
Db 2263 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2322
Qy 2517 TTGCAGAGGTTGACTTTTCTCGCAGTGCAGTGTAAAGTGAATGGTCAATGACCAAGGCTCAG 2576
Db 2323 TTGCAGAGGTTGACTTTTCTCGCAGTGCAGTGTAAAGTGAATGGTCAATGACCAAGGCTCAG 2382
Qy 2577 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAA 2636
Db 2383 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAA 2442
Qy 2637 ACCATAGCGGAATGTGACAAAAGAAAGCCCCATCTCTGATTTGCCACTGGAAAGC 2696
Db 2443 ACCATAGCGGAAATGTGACAAAAGAAAGCCCCATCTCTGATTTGCCACTGGAAAGC 2502
Qy 2697 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2756
Db 2503 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2562
Qy 2757 AATGCTGAATCATCATAGCGGTTACTGAAGTGTACTTGGAGAGAGCTGACGCAATTAC 2816
Db 2563 AATGCTGAATCATCATAGCGGTTACTGAAGTGTACTTGGAGAGAGCTGACGCAATTAC 2622
Qy 2817 ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTC 2876
Db 2623 ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTC 2682
Qy 2877 GACGAGAGAGTCAATTCAGACACAAATTTGGCATCTTCTCATAGTACAAAATCTCTGG 2936
Db 2683 GACGAGAGAGTCAATTCAGACACAAATTTGGCATCTTCTCATAGTACAAAATCTCTGG 2742
Qy 2937 AGGCACTAAAGAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC 2996
Db 2743 AGATACAGAGAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC 2802
Qy 2997 AAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTCACCGGAT 3056
Db 2803 AAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTCACCGGAT 2862
Qy 3057 GTTATGCAAGAACGTCAGATATGTTCCCTGAGCAATTTGACAGAGCTCATTTTCACAA 3116
Db 2863 GTTATGCAAGAACGTCAGATATGTTCCCTGAGCAATTTGACAGAGCTCATTTTCACAA 2922
Qy 3117 GTGCGCCATTCAGGAGATTCAGTGTTCCTTTCTTTATTTTATTTATCTCATGAGTGCA 3176
Db 2923 GTGCGCCATTCAGGAGATTCAGTGTTCCTTTCTTTATTTTATTTATCTCATGAGTGCA 2982
Qy 3177 GTGCGCCATTCAGTATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTG 3236
Db 2983 GTGCGCCATTCAGTATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTG 3042
Qy 3237 TTGTCTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCACGAACTGCCGTTAAGTTTG 3296
Db 3043 TTGTCTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCACGAACTGCCGTTAAGTTTG 3102
Qy 3297 CAGGATTTGACAGGTCCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3356
Db 3103 CAGGATTTGACAGGTCCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3162
Qy 3357 ATCACGAGCTAAATAATATTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3416
Db 3163 ATCACGAGCTAAATAATATTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3222
Qy 3417 CCGGTCTCTAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCCAAAAGCA 3476
Db 3223 CCGGTCTCTAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCCAAAAGCA 3282
Qy 3477 TATAAGGACAAAACAAATATAGGTTTGAATCATGCGGAGAGAAATCGCTTTTAAA 3536
Db 3283 TATAAGGACAAAACAAATATAGGTTTGAATCATGCGGAGAGAAATCGCTTTTAAA 3342

Qy 3537 ATGATTCTACCAAGTTTTTCTCATGTGGTGGCCAGTTGGATGACATAAAGAAAAACCCCT 3596
Db 3343 ATGATTCTACCAAGTTTTTCTCATGTGGTGGCCAGTTGGATGACATAAAGAAAAACCCCT 3402
Qy 3597 AGGAAAGTTTGTGGCTGAATGACAAACATTTGACCAACATCAATAAGATGCTCAGACAGTG 3656
Db 3403 AGGAAAGTTTGTGGCTGAATGACAAACATTTGACCAACATCAATAAGATGCTCAGACAGTG 3462
Qy 3657 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG 3716
Db 3463 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG 3522
Qy 3717 CCAAGAGAGATTCGAAACCGTTTCTTCTCATATGATGAGCTGCAGGAATGGAGGGCTTAT 3776
Db 3523 CCAAGAGAGATTCGAAACCGTTTCTTCTCATATGATGAGCTGCAGGAATGGAGGGCTTAT 3582
Qy 3777 CGAGCAAAATTTGAAGT 3792
Db 3583 CGAGCAAAATTTGAAGT 3598

Search completed: November 22, 2004, 16:51:27
Job time : 2633 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 22:56:56 ; Search time 15702 Seconds
(without alignments)
12988.994 Million cell updates/sec

Title: US-10-023-888-3

Perfect score: 5597
Sequence: 1 cggagcgcagggcgctcg.....aaaagttaatttggaaa 5597

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 segs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsl1.*
9: gb_gsl2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	856.2	15.3	934	5	BQ719691 AGENCOURT
2	829.6	14.8	1016	4	BM544697 AGENCOURT
3	813	14.5	994	4	BM476404 AGENCOURT
4	812	14.5	981	5	BQ053143 AGENCOURT
5	804.8	14.4	1118	4	BM452210 AGENCOURT
6	794.8	14.2	855	6	CD657400 AGENCOURT
7	771.8	13.8	1061	4	BM450679 AGENCOURT
8	770	13.8	854	5	BU171738 AGENCOURT
9	751.8	13.4	867	6	CD657425 AGENCOURT
10	746.8	13.3	929	5	BQ720582 AGENCOURT
11	743.8	13.3	783	4	BQ214021 AGENCOURT
12	733.8	13.1	819	6	CD652380 AGENCOURT
13	728.4	13.0	826	5	BU933148 AGENCOURT
14	715.2	12.8	773	4	BG196325 AGENCOURT
15	710.8	12.7	795	4	BG194321 AGENCOURT
16	694.2	12.4	899	6	CD244899 AGENCOURT
17	692	12.4	780	1	AU134813 AGENCOURT
18	680.4	12.2	814	6	CD657278 AGENCOURT
19	671	12.0	984	4	BG250072 AGENCOURT
20	665.2	11.9	778	6	CD643692 AGENCOURT
21	664.8	11.9	705	5	EX095637 AGENCOURT
22	644.8	11.5	686	5	BQ008614 UI-H-ED1-
23	643.8	11.5	792	4	BG200937 RST20259
24	637.6	11.4	682	6	CA438470 UI-H-DT1-

25	636.8	11.4	794	6	CD656815
26	623.8	11.1	852	4	BG199915
27	622.6	11.1	630	7	CN426009 170006000
28	620.6	11.1	842	4	BG207173 RST26640
29	617.8	11.0	773	7	CF406980 CH3#045_F
30	617.2	11.0	634	5	EX645182 DKF2D781D
31	608	10.9	670	1	AU135455 AU135455
32	605.6	10.8	789	7	CN525214 UI-M-GHO-
33	600	10.7	783	7	CK781711 UI-M-GHO-
34	599.8	10.7	613	4	BG163659 602338839
35	599.4	10.7	791	6	CB518925 UI-M-GHO-
36	598.4	10.7	639	7	CK002110 AGENCOURT
37	596	10.6	793	6	CB527476 UI-M-FY0-
38	593.2	10.6	776	7	CF406981 CH3#045_F
39	589.2	10.5	616	4	BM790504 K-EST0070
40	588.6	10.5	819	6	CA324819 UI-M-FY0-
41	584.6	10.4	731	5	BU709362 UI-M-EVO-
42	575	10.3	663	7	CO261199 4133319_B
43	573.2	10.2	612	4	BG577274 602595465
44	573	10.2	782	5	BU451401 603215493
45	568.8	10.2	1037	5	BQ226717 AGENCOURT

ALIGNMENTS

RESULT 1

BQ719691

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ719691 934 bp mRNA linear EST 16-JUL-2002
AGENCOURT_8304671 Lupski_sympathetic_trunk Homo sapiens CDNA clone
IMAGE:6193737 5', mRNA sequence.

BQ719691 GI:21858588
EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 934)

NIH-MGC http://mgc.ncl.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13597 row: n column: 10

High quality sequence stop: 626.

Location/Qualifiers

1..934

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6193737"

/sex="male"

/tissue_type="sympathetic trunk"

/dev_stage="adult, 16 yr"

/lab_host="DH10B"

/clone_lib="Lupski_sympathetic trunk"

/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:

NotI; Site_2: SalI; CDNA made by oligo-dT priming.

Directionally cloned using the following adaptors:

5'-TCGACCCACGCTCCG-3' and

5'-GACTAGTCTAGATCCGAGCGCCCGCTT(15)-3'. Size selected >

1 kb for average insert length 1.9 kb. This is a primary

library, non-amplified. Library constructed by Life

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor

College of Medicine); available through Life

ORIGIN		Technologies."		ORIGIN			
Query Match		15.3%; Score 856.2; DB 5; Length 934;		Query Match			
Best Local Similarity		97.6%; Pred. No. 3.4e-127; Indels 4; Gaps 4;		Best Local Similarity			
Matches 911; Conservative 0; Mismatches 18;				Matches 863; Conservative 0; Mismatches 19; Indels 2; Gaps 2;			
QY	3313	TGGAACACATGCTCTAATAAATGCTCAAAAATGCTTCTGCTGATATCAAGCAGCTAAATA	3372	QY	193	CCTATACCTGCTCTCCACAGGATGAGTATGGCTCTACGTGTGCTTCTGGGGTCTGTTTCA	252
DB	1	TGGAACACATGCTCTAATAAATGCTCAAAAATGCTTCTGCTGATATCAAGCAGCTAAATA	60	DB	1	CCTATACCTGCTCTCCACAGGATGAGTATGGCTCTACGTGTGCTTCTGGGGTCTGTTTCA	60
QY	3373	ATATTCACCACTCAGGATCTCTATGATCCCACTGCGACCGTCACTAAAGTC	3432	QY	253	CCATCGTCTCCGCTTCCAGTTCGAGAGGTTCTGGAATGAGCGGAGATCAATACC	312
DB	61	ATATTCACCACTCAGGATCTCTATGATCCCACTGCGACCGTCACTAAAGTC	120	DB	61	CCATCGTCTCCGCTTCCAGTTCGAGAGGTTCTGGAATGAGCGGAGATCAATACC	120
QY	3433	TAGTAACAACTGTAAACCACTGTAATGATCCCACTGCGACCGTCACTAAAGTC	3492	QY	313	ATGTTTGTGTTGATTCCTATAGACAAATATGCTGGAAGTCTCTTTTTCAGAAATCGGCTTT	372
DB	121	TAGTAACAACTGTAAACCACTGTAATGATCCCACTGCGACCGTCACTAAAGTC	180	DB	121	ATGTTTGTGTTGATTCCTATAGACAAATATGCTGGAAGTCTCTTTTTCAGAAATCGGCTTT	180
QY	3493	AATATAGTTTGAATCATGCGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAG	3552	QY	373	GTCTGCCCATCCGATTCACGTTGTTTACACCTGGGTGAATGGCACAGATCTTGAACCTAC	432
DB	181	AATATAGTTTGAATCATGCGAGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAG	240	DB	181	GTCTGCCCATCCGATTCACGTTGTTTACACCTGGGTGAATGGCACAGATCTTGAACCTAC	240
QY	3553	TTTCTCATGTTGTTGGCCAGTTGGATGACATAGAAAAACCTTAGGAAGTTTGTGTC	3612	QY	433	TGAAGGAATCAACAGAGTCAAGACAGATGAGAGGAGGAGCAGAAAGCAATGAGAGAA	492
DB	241	TTTCTCATGTTGTTGGCCAGTTGGATGACATAGAAAAACCTTAGGAAGTTTGTGTC	300	DB	241	TGAAGGAATCAACAGAGTCAAGACAGATGAGAGGAGGAGCAGAAAGCAATGAGAGAA	300
QY	3613	TGAATGACAACTGACCAAACTAATAAGATGCTCAGACAGTGAAGGCTTTCTCAGGG	3672	QY	493	TCCTTGGGAAAAACACAAACCGAAGCTTACTAAGAGAGTGAAGACAGTGTAGAGTTTGC	552
DB	301	TGAATGACAACTGACCAAACTAATAAGATGCTCAGACAGTGAAGGCTTTCTCAGGG	360	DB	301	TCCTTGGGAAAAACACAAACCGAAGCTTACTAAGAGAGTGAAGACAGTGTAGAGTTTGC	360
QY	3673	ACTCTATGATCATGCTCCCATATCTCCCAATTTGAATGACGCAAGAGATTCGAA	3732				
DB	361	ACTCTATGATCATGCTCCCATATCTCCCAATTTGAATGACGCAAGAGATTCGAA	420				
QY	3733	ACCGTTTCTTCATATGATGATGCTGAGGAATGAGGGGTTATCGAGACAAATCGAAT	3792				
DB	421	ACCGTTTCTTCATATGATGATGCTGAGGAATGAGGGGTTATCGAGACAAATCGAAT	480				
QY	3793	TTTGAGCCCATGTTGATGACATGATGATGATGATGATGATGATGATGATGATGATG	3852				
DB	481	TTTGAGCCCATGTTGATGACATGATGATGATGATGATGATGATGATGATGATGATG	540				
QY	3853	AGCAGTTAATGCACTTAAAGCGGAAGATATTTCCAGAGGAGGATACACAAAGAGCTA	3912				
DB	541	AGCAGTTAATGCACTTAAAGCGGAAGATATTTCCAGAGGAGGATACACAAAGAGCTA	600				
QY	3913	GTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTACCTCAGCATTTA	3972				
DB	601	GTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTACCTCAGCATTTA	560				
QY	3973	CTGAGCATTTTAAACCTCAGCTTACAGAGATGCTTTTGTGATGCTGATGCTTACAGATT	4032				
DB	661	CTGAGCATTTTAAACCTCAGCTTACAGAGATGCTTTTGTGATGCTGATGCTTACAGATT	720				
QY	4033	GGCCCGAAGAGGAAATATCCAGTACCATGCTG- TTTTGTGCAATGAATATAGCCCAAC	4090				
DB	721	GGCCCGAAGAGGAAATATCCAGTACCATGCTG- TTTTGTGCAATGAATATAGCCCAAC	780				
QY	4091	TGACTAGGAATATTTAAACCAACCCACTGAAACTGTTGTTGTCGAGCAGCTCTGAACTGA	4150				
DB	781	GGACAGGGAATATTTTAAACCAACCCACTGAAACTGTTGTTGTCGAGCAGCTCTGAACTGA	840				
QY	4151	TTTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCCTTTTTTATAAAGGCTCTACTG	4209				
DB	841	TTTTTACTTTTAAAGAAATTTGCTCATGACCTGTCATCCTTTTTTATAAAGGCTCTACTG	900				
QY	4210	ACAAGACACAGCT-GTTAATTTCCACAGCAAT	4241				
DB	901	ACAAGACACAGCTGTTAATTTCCACAGCAAT	933				

	553	TAAACACTGCATTAAAGTGGCAATGCTTGATCGGACCAGCCCTGCCAGCCACAATCA	611
QY			
Db	361	TAAACACTGCATTAAAGTGGCAATGCTTGATCGGACCAGCCCTGCCAGCCACAATCA	420
QY			
Db	613	CCTCGAAGGACGTGCATCTCTTTATCCTTCCTTTTCATCTGCCAGTAGCATTTTCAATG	672
QY			
Db	421	CCTCGAAGGACGTGCATCTCTTTATCCTTCCTTTTCATCTGCCAGTAGCATTTTCAATG	480
QY			
Db	673	TTGCAAAAACAAAAAACCCCTTCTACCAATGTCCTAGTGTGTTTGTGAAGTAAGG	732
QY			
Db	481	TTGCAAAAACAAAAAACCCCTTCTACCAATGTCCTAGTGTGTTTGTGAAGTAAGG	540
QY			
Db	733	ATGTTGAAGATCCCACACTCTGAGCTGCTTAAAGGAATAGCACAGACAGTATGGAGG	792
QY			
Db	541	ATGTTGAAGATCCCACACTCTGAGCTGCTTAAAGGAATAGCACAGACAGTATGGAGG	600
QY			
Db	793	GGTACTTGACAAACAGATAAGAAGTCCCTGATAGTGTAAATGCAAGATTTGGCTTTCC	852
QY			
Db	601	GCTACTTGACAAACAGATAAGAAGTCCCTGATAGTGTAAATGCAAGATTTGGCTTTCC	660
QY			
Db	853	TGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAAAAACAAAATTCGCAGAAA	912
QY			
Db	661	TGAGTGGATTTCACCAACATTCAGGAAACAAATCAACTAAAAACAAAATTCGCAGAAA	720
QY			
Db	913	ATCTTTCTCTAAAGTCAAACGTGTGACAGTGTGATTCAGAGGCCAGTGTAGCGCTTCAA	972
QY			
Db	721	ATCTTTCTCTAAAGTCAAACGTGTGACAGTGTGATTCAGAGGCCAGTGTAGCGCTTCAA	780
QY			
Db	973	AAGTGAATAACCCCAAGGATTTCAAGAAATGAATAAG-CAAACTAAGAAAGCAATGACC	1033
QY			
Db	781	AACTGAATAACCCCAAGGATTTCAAGAAATGAATAAGCCAACTAAGAAAGCAATGACC	840
QY			
Db	1032	ATTG-ATGGAAGAAAGCACTGACATAAGTCTGCAATATTTATTAT 1074	
QY			
Db	841	ATTGATGGAAGAAAGCACTGACATAAGTCTGCAATATTTATTAT 884	
QY			
Db			
RESULT 3			
BM476404			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NH-MGC http://mgc.nci.nih.gov/.			
1 bases 1 to 994			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999).			
Contact: Robert Strausberg, Ph.D.			
Email: cgapbs-r@mail.nih.gov			
Tissue Procurement: Lou Staudt			
cDNA Library Preparation: Life Technologies, Inc.			
cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Agencourt Bioscience Corporation			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLAM12270 row: d column: 04			
High quality sequence stop: 676.			
Location/Qualifiers			
1..994			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5553363"			
/tissue_type="lymphoma, cell line"			
/lab_host="DH10B (phage-resistant)"			
/clone_lib="NIH MGC 85"			
FEATURES			
source			

Db 901 AACCTGGATCTCCACATATTTCTTTCCCAAAAGGGGATACATCTTTTG 949

RESULT 4
BQ053143
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ053143 981 bp mRNA linear EST 29-MAR-2002
AGENCOURT 6821868 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5934740
5' mRNA sequence.
BQ053143
EST.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 981)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2121 row: f column: 21
High quality sequence stop: 575.
Location/Qualifiers
1. 981
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5934740"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 106"
/note="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2:
ECORI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Query Match 14.58; Score 812; DB 5; Length 981;
Best Local Similarity 95.48; Pred. No. 4e-120;
Matches 867; Conservative 0; Mismatches 38; Indels 4; Gaps 3;

Qy 3315 GAACACATGCTATAAATGCTCAAAAATGCTTCCTGCTGATATACGCGAGCTAAATAAT 3374
Db 1 GAACACATGCTATAAATGCTCAAAAATGCTTCCTGCTGATATACGCGAGCTAAATAAT 60

Qy 3375 ATTCCACCACTCGGAATCCTACTATGATCCCACTGCGCCCGGTCACTAAAGTCTA 3434
Db 61 ATTCCACCACTCGGAATCCTACTATGATCCCACTGCGCCCGGTCACTAAAGTCTA 120

Qy 3435 GTACCAAACTGTAACACAGTAACCTGACAAAATCCAAAGCATATAGGACAAAACAAA 3494
Db 121 GTACCAAACTGTAACACAGTAACCTGACAAAATCCAAAGCATATAGGACAAAACAAA 180

Qy 3495 TATAGTTTGAATCATGCGAGGAAGAGAAATCGCTTTTAAATGATTCGTACCAACGTT 3554
Db 181 TATAGTTTGAATCATGCGAGGAAGAGAAATCGCTTTTAAATGATTCGTACCAACGTT 240

Qy 3555 TCTCATGTGTTGCCAGTTGGATGACATAGAAAACCCCTAGGAAGTTTGTTCCTG 3614
Db 241 TCTCATGTGTTGCCAGTTGGATGACATAGAAAACCCCTAGGAAGTTTGTTCCTG 300

Qy 3615 AATGACAACATTGACCACATATATAAGATGCTTCAGACAGTGAAGGCTGTTCTCAGGGAC 3674
Db 301 AATGACAACATTGACCACATATATAAGATGCTTCAGACAGTGAAGGCTGTTCTCAGGGAC 360

Qy 3675 TTCATGAATCCATGTTTCCCATACCTTCCCAATTTGAATTCGCAAGAGAGTATCGAAAC 3734
Db 361 TTCATGAATCCATGTTTCCCATACCTTCCCAATTTGAATTCGCAAGAGAGTATCGAAAC 420

Qy 3735 CGTTTCCTTCATATGATGAGCTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTTT 3794
Db 421 CGTTTCCTTCATATGATGAGCTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTTT 480

Qy 3795 TGGACCCATGTTGTTACTAGCAACATTGATTTACTATTTACTATTTCTCATTTTGTCTGAG 3854
Db 481 TGGACCCATGTTGTTACTAGCAACATTGATTTACTATTTACTATTTCTCATTTTGTCTGAG 540

Qy 3855 CAGTTAATTCGACTTAAGCGGAAGATATTTCCAGAGAGGAGTACACAAAGAGCTAGT 3914
Db 541 CAGTTAATTCGACTTAAGCGGAAGATATTTCCAGAGAGGAGTACACAAAGAGCTAGT 600

Qy 3915 CCCAATCGAATCAGAGTATAGAGATCTTTCATTTTGAACCATCTACCTCAGCATTTTACT 3974
Db 601 CCCAATCGAATCAGAGTATAGAGATCTTTCATTTTGAACCATCTACCTCAGCATTTTACT 660

Qy 3975 GAGCATTTTAAACTCAGCTTCAGAGATGTCCTTTGTGATGATGCTTAGCAG-TTTG 4033
Db 661 GAGCATTTTAAACTCAGCTTCAGAGATGTCCTTTGTGATGATGCTTAGCAG-TTTG 720

Qy 4034 GCCCGAAGAGGAAATATCCAGTACCATGCTGTTTGTGCGATGAATAGCCCACTGA 4093
Db 721 CCCCGAAGAGGAAATATCCAGTACCATGCTGTTTGTGCGATGAATAGCCCACTGA 780

Qy 4094 CTAGGAATTTTAAACCAACCCACTG-AAACTTGTGTGCGAGCAGCTCTGAACCTGATT 4152
Db 781 CCAGGAATTTTAAACCAACCCACTGAAACCACTGTTGTTTGCAGCAGCTCTGAACCTGATT 840

Qy 4153 TTACTTTTAAAGATTTT-GCTCATGACCTGTCATCTTTTATAAAAGGCTCACTGA 4210
Db 841 TTACTTTTAAAGATTTTGTCTTCAAGCCGTCGTCATCTTTTATAAAAGGCTCA 900

Qy 4211 CAAGAGACA 4219
Db 901 CCTGACAAA 909

RESULT 5
BQ452210
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ452210 1118 bp mRNA linear EST 05-FEB-2002
AGENCOURT 6386314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526571
5' mRNA sequence.
BQ452210
BQ452210.1 GI:18501250
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1118)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/BCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M12200 row: g column: 20
High quality sequence stop: 717.
Location/Qualifiers

FEATURES

```
source
1. 1118
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5526571"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-Sport6; Site:1: NotI,
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match
Best Local Similarity 14.4%; Score 804.8; DB 4; Length 1118;
Matches 885; Conservative 0; Mismatches 46; Indels 15; Gaps 4;

3946 TTTGAAACCACTCTACCTCAGCATTTACTGAGCAATTTTAAACCTCAGCTTCACAGAGATG 4005
Db |||
57 TTTGAAACCACTACCTCAGCATTTACTGAGCAATTTTAAACCTCAGCTTCACAGAGATG 116
QY |||
4006 TCTTTGTGATGATGCTTAGCAGTTTGGCCGGAAGAGAAATATCCAGTACCATGCT 4065
Db |||
117 TCTTTGTGATGATGCTTAGCAGTTTGGCCGGAAGAGAAATATCCAGTACCATGCT 176
QY |||
4066 GTTTTGTGCGATGAATATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGAAACT 4125
Db |||
177 GTTTTGTGCGATGAATATAGCCCACTGACAGGAATTTTAAACCAACCCACTGAAACT 236
QY |||
4126 TGTGTGTCGACAGCTCTGAATGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCA 4185
Db |||
237 TGTGTGTCGACAGCTCTGAATGATTTTACTTTTAAAGAAATTTGCTCATGACCTGCA 296
QY |||
4186 TCCCTTTTATAAAGGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCAT 4245
Db |||
297 TCCCTTTTATAAAGGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCAT 356
QY |||
4246 GCAGACTAACTTTATTAGGAAGACCTATGCGAGCTGGAGTATTGCTTAAGAGCTCCA 4305
Db |||
357 GCAGACTAACTTTATTAGGAAGACCTATGCGAGCTGGAGTATTGCTTAAGAGCTCCA 416
QY |||
4306 GTCTTTGCAATCCAAAGCCTTTTCTAAAGTTTGGCACTTTTCTTTTCTTCCATT 4365
Db |||
417 GTCTTTGCAATCCAAAGCCTTTTCTAAAGTTTGGCACTTTTCTTTTCTTCCATT 476
QY |||
4366 TTTAAGTAGTTACTAAGTTAAGTTATTTCTCTCTCTGAGTATAAGCAATTTGGGATG 4425
Db |||
477 TTTAAGTAGTTACTAAGTTAAGTTATTTCTCTCTCTGAGTATAAGCAATTTGGGATG 536
QY |||
4426 CTAAACCTATTTTATAGATGTTATTTAAATAATGCAAGCAATATCACTCTTATTGCAA 4485
Db |||
537 CTAAACCTATTTTATAGATGTTATTTAAATAATGCAAGCAATATCACTCTTATTGCAA 596
QY |||
4486 TACCTAAATATGAGTTTATTAATATTTAAGCTGTAATGCTTAAACCACTAAGT 4545
Db |||
597 TACCTAAATATGAGTTTATTAATATTTAAGCTGTAATGCTTAAACCACTAAGT 656
QY |||
4546 CTGAAGAGCTCAATGATGATCATCTGAAATGCTTTGTAATTTAGCTTCAGCCCTAAG 4605
Db |||
657 CTGAAGAGCTCAATGATGATCATCTGAAATGCTTTGTAATTTAGCTTCAGCCCTAAG 716
QY |||
4606 AATGCTATGATTTACGTGAGGCTTAATTTCAACAGCTAGAGTATGATCTACTACCA 4665
Db |||
717 AATGCTATGATTTACGTGAGGCTTAATTTCAACAGCTAGAGTATGATCTACTACCA 776
QY |||
4666 GATGTAATTTATGTTTGGAAATGATACATATTTCAACAGAGTGCCTCATTTAGAATGA 4725
Db |||
777 GATGTAATTTATGTTTGGAAATGATACATATTTCAACAGAGTGCCTCATTTAGAATGA 836
QY |||
4726 GTAGTGCTGATGGCACTGGCACT---ACAGTGGTCTTGTTTAACTACTATTTGTA- 4781
Db |||
837 GTAGTGCTGATGGCACTGGCACTAATTAACAGGGGNGNCTTGGNTTAATACTACTTGGNA 896
```

```
QY 4782 -TATTCAGTAGCTATCTCTCTC-----AGTTGGTTTTTATAGACAGAGCCAGC 4832
Db 897 TATTCAGTAGCTATCTCTCTCAGTTGGGTTTTTGAATAGAAACAGAGGGCCAGC 956
QY 4833 AACTTTTC--TTTGTAAAAGCTGGTTAGTAATATATTCAGGCCA 4876
Db 957 AACTTTTCCTTTTGTAAAAGCTGGTTAAAGAAATTTATGGCAA 1002
```

```
RESULT 6
CD657400
LOCUS
DEFINITION
AGENCOURT_14538430 NIA Human H1 Embryonic Stem Cell cDNA Library
(long) Homo sapiens cDNA clone IMAGE:30420087 5', mRNA sequence.
CD657400
CD657400.1 GI:31899587
EST.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 855)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
cDNA Library Preparation: Yulan Piao and Minoru Ko
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC c lone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM495 row: 1 column: 16
High quality sequence stop: 684.
Location/Qualifiers
```

FEATURES

```
source
1. 855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30420087"
/tissue_type="Embryonic Stem cells"
/cell_line="WA01"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library
(long)"
/note="Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI;
This is a long-transcript enriched cDNA library (Genome
Res 11: 1553-1558 (2001). [PMID: 11544199]) from WA01
cell line. Undifferentiated human ES cell line WA01/H1
was obtained from WiCell Research Institute, Inc.,
Madison, WI, cultured according to their instructions, on
MEF feeders. They formed round colonies with defined edges
and were positive for alkaline phosphatase, SSEA-4, OCT3,
OCT4, REX1, UTF, TERT, SOX2, CX43 and CX45. They are
negative for GATA2, GATA3, PDX1, NCAM, MSX1, FLK3, SSEA-1,
TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days
after plating), the ES cells from 4 X 6cm dishes were
treated with 1 mg/ml collagenase, type IV
(invitrogen/Gibco) for 5-10 min and gently scraped off
with 5 ml pipette. RNA was purified with Trizol Reagent
from invitrogen. Protocol ref: Genome Res 11: 1553-1558
(2001). [PMID:11544199]) Double-stranded cDNAs were
synthesized with an oligo(dT) primer [invitrogen:
5'-pGACTAGTCTAGATCGGAGCGCCCTTTTCTTTT-3'] from
3.4g of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lone-linker LL-Sal4, purified by phenol/chloroform
```


extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of PCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN		Query Match	Score	794.8;	DB 6;	Length	855;
		Best Local Similarity	98.1%;	Pred. No. 2.3e-117;			
		Matches	810;	Conservative	0;	Mismatches	15;
						Indels	1;
						Gaps	
QY	2616	GTGAAACTCACCACCCCAAAACCATAGCGGGAATGTGACAAAGAAAGCCCATCT	2675				
DB	12	GTGAAACTCACCACCCCAAAACCATAGCGGGAATGTGACAAAGAAAGCCCATCT	71				
QY	2676	CTGATTCTTCCACTGGAAGCCAGATGACAAAGAAAGAAATACACAGGAAAGAAAA	2735				
DB	72	CTGATTCTTCCACTGGAAGCCAGATGACAAAGAAAGAAATACACAGGAAAGAAAA	131				
QY	2736	GAGAACAGTAGAATGGAGGAAATGCTGAAATACATAGCGGCTTACTGAAAGTTACTT	2795				
DB	132	GAGAACAGTAGAATGGAGGAAATGCTGAAATACATAGCGGCTTACTGAAAGTTACTT	191				
QY	2796	GGAAGAAAGCTGACGATACACAGATGATCTGGCTTTTGGCATGGGAGAAAAA	2855				
DB	192	GGAAGAAAGCTGACGATACACAGATGATCTGGCTTTTGGCATGGGAGAAAAA	251				
QY	2856	AAGTATTTCCTCAAGATCTTCGACGAGAAAGAGTCATTGGAACACAAATGGCATCTTC	2915				
DB	252	AAGTATTTCCTCAAGATCTTCGACGAGAAAGAGTCATTGGAACACAAATGGCATCTTC	311				
QY	2916	ACTGATGACAAATATCTGGGACCACTAAAGATACATTCGACATCCCTCAGATAT	2975				
DB	312	ACTGATGACAAATATCTGGGACCACTAAAGATACATTCGACATCCCTCAGATAT	371				
QY	2976	GTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCTGTCTCATG	3035				
DB	372	GTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCTGTCTCATG	431				
QY	3036	CCTCAGATGATGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATT	3095				
DB	432	CCTCAGATGATGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATT	491				
QY	3096	GACAGAGCTCATTTCAAGTCGCGCATCTGAGGATATGAGATTTGGCTTCTCTTAT	3155				
DB	492	GACAGAGCTCATTTCAAGTCGCGCATCTGAGGATATGAGATTTGGCTTCTCTTAT	551				
QY	3156	TTTTATTATCTCATGAGTCAGTCAGCCACTGAATATATCTCAAGTCTTTGATGAATT	3215				
DB	552	TTTTATTATCTCATGAGTCAGTCAGCCACTGAATATATCTCAAGTCTTTGATGAATT	611				
QY	3216	GATCAGATCAATCTGGTCTTGTCTGACAGAGAAATCCGACACTGGCTACCAGAATT	3275				
DB	612	GATCAGATCAATCTGGTCTTGTCTGACAGAGAAATCCGACACTGGCTACCAGAATT	671				
QY	3276	CACGAATCTGCGTTAAGTTTTCAGGATTTGACAGGTCGGAACACATGCTTAATAATTGC	3335				
DB	672	CACGAATCTGCGTTAAGTTTTCAGGATTTGACAGGTCGGAACACATGCTTAATAATTGC	731				
QY	3336	TCAAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCAGGATCC	3395				
DB	732	TCAAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCAGGATCC	791				
QY	3396	TACTATGATCCCAACCTGCCACCGGTCACTAAAAAGTCTAGTAACAA	3441				
DB	792	TACTATGATCCCAACCTGCCACCGGTCACTAAAAAGTCTAGTAACAA	836				

BM450679
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BM450679
AGENCOURT_6394543 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494379
5', mRNA sequence.
BM450679
BM450679.1 GI:18499719
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1061)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12119 row: j column: 12
High quality sequence stop: 637.
Location/Qualifiers
1..1061
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5494379"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN		Query Match	Score	771.8;	DB 4;	Length	1061;
		Best Local Similarity	95.9%;	Pred. No. 1.1e-113;			
		Matches	826;	Conservative	0;	Mismatches	27;
						Indels	8;
						Gaps	3;
QY	640	CTTCTTTTCATCTGCCAGTGCACATTTTCATGTTGCAAAACCAAAACCCCTCTACCA	699				
DB	25	CTTCTTTTCATCTGCCAGTGCACATTTTCATGTTGCAAAACCAAAACCCCTCTACCA	84				
QY	700	ATGCTCAGTTGTTGTTTTCACAGTACTAAGAGTGTGAAGATGCCCACTCTGCACTGC	759				
DB	85	ATGCTCAGTTGTTGTTTTCACAGTACTAAGAGTGTGAAGATGCCCACTCTGCACTGC	144				
QY	760	TTAAAGAAATAGCAGACAGACAGATGAGAGGGGTACTTCACACAGATTAAGAGTCC	819				
DB	145	TTAAAGAAATAGCAGACAGACAGATGAGAGGGGTACTTCACACAGATTAAGAGTCC	204				
QY	820	CTGGATTAGTCTAATCAAGATTTGGCTTTCTCTGAGTGGATTTCCCAACATTCAGG	879				
DB	205	CTGGATTAGTCTAATCAAGATTTGGCTTTCTCTGAGTGGATTTCCCAACATTCAGG	264				
QY	880	AAACAAATCAACTAAAAACAAAATTGCCAGAAAATCTTTTCTCTAAAGTCAAACTGTTC	939				
DB	265	AAACAAATCAACTAAAAACAAAATTGCCAGAAAATCTTTTCTCTAAAGTCAAACTGTTC	324				
QY	940	AGTTGATTCAGAGCCAGCTGTGCGCTTCTTAAACCTGAATAACCCCAAGGATTTTCAAG	999				
DB	325	AGTTGATTCAGAGCCAGCTGTGCGCTTCTTAAACCTGAATAACCCCAAGGATTTTCAAG	384				
QY	1000	AATTGAATAAGCAAACTAAGAGAAACATGACCATTTGATGGAAAAAGAACTGACCAATGTC	1059				
DB	385	AATTGAATAAGCAAACTAAGAGAAACATGACCATTTGATGGAAAAAGAACTGACCAATGTC	444				

QY 1060 CTGCATATTTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCT 1119
|||||
Db 445 CTGCATATTTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCT 504
|||||
QY 1120 CTGCCAGTCGTTTTGAAGATAACGAAGAATCTGAGTACTCTATTCGATCTATCGAGAGGC 1179
|||||
Db 505 CTGCCAGTCGTTTTGAAGATAACGAAGAATCTGAGTACTCTATTCGATCTATCGAGAGGC 564
|||||
QY 1180 ATGCACCATGGTTCGGAATATTTTCAATGTGTCACCAAGCGGCAGATCCATCTGGGTGA 1239
|||||
Db 565 ATGCACCATGGTTCGGAATATTTTCAATGTGTCACCAAGCGGCAGATCCATCTGGGTGA 624
|||||
QY 1240 ACCTTGACAACTCTCGAGTGAACAATAGTAACACACACAGGATGTTTTTCGAAATTTGAGCC 1299
|||||
Db 625 ACCTTGACAACTCTCGAGTGAACAATAGTAACACACAGGATGTTTTTCGAAATTTGAGCC 684
|||||
QY 1300 ACTTGCCTACCTTTAGTTTACCTGCTATTTGAAAGTCAATTCATCGCATCGAAGGGCTGT 1359
|||||
Db 685 ACTTGCCTACCTTTAGTTTACCTGCTATTTGAAAGTCAATTCATCGCATCGAAGGGCTGT 744
|||||
QY 1360 CCCAGAGTTTATTTTACCTAATGATGATGTCATGTTGGG---AAGGATCTCTGGCCAG 1416
|||||
Db 745 CCCAGAGTTTATTTTACCTAATGATGATGTCATGTTGGG---AAGGATCTCTGGCCAG 804
|||||
QY 1417 ATGATTTTTACAGTCAC--TCCAAAGGCCAGAGGTTTATTTGACATGGCC---TGTGCC 1471
|||||
Db 805 ATGATTTTTACAGTCACCTCCAAAGGCCAGAGGTTTATTTGACATGGCCCTGTGCCCA 864
|||||
QY 1472 AAACCTGTCGCGAGGGCTGCC 1492
|||||
Db 865 AAACCTGTCGCGAGGGCTGCC 885
|||||

RESULT 8
BUI171738
LOCUS
DEFINITION AGENCOURT_7936624 NIH_MGC_72 Homo sapiens cdna clone IMAGE:6147890
5', mRNA sequence.
BUI171738
VERSION BUI171738.1 GI:22685722
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 854)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC/DCTP/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAMI3478 row: h column: 03
High quality sequence stop: 624.
Location/Qualifiers
1. 854
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6147890"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: Skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."

FEATURES
source

ORIGIN
Query Match 13.8%; Score 770; DB 5; Length 854;
Best Local Similarity 99.0%; Pred. No. 2.2e-113;
Matches 806; Conservative 0; Mismatches 5; Indels 3; Gaps 3;
QY 3099 AAGACGTCATTTTACAAAAGTGGCCCATCTCGAGGATATCGAGTTGCTCTCTTATTTT 3158
Db 1 AAGACGTCATTTTACAAAAGTGGCCCATCTCGAGGATATCGAGTTGCTCTCTTATTTT 60
QY 3159 TATTATCTCATGAGTGCAGTGCAGCCACATGATATATCTCAAGTCTTTTATGATCAAGTTGAT 3218
Db 61 TATTATCTCATGAGTGCAGTGCAGCCACATGATATATCTCAAGTCTTTTATGATCAAGTTGAT 120
QY 3219 ACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACTCTGGTACCAAGATTCTAC 3278
Db 121 ACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACTCTGGTACCAAGATTCTAC 180
QY 3279 GAATCCCGTTAAGTTTTCAGGATTTGACAGGCTCTGGAACACATGCTAATAATTGCTCA 3338
Db 181 GAATCCCGTTAAGTTTTCAGGATTTGACAGGCTCTGGAACACATGCTAATAATTGCTCA 240
QY 3339 AAAATGCTTCTCTGCTGATATCAGCAGCTATAATATATTTCCACCAACTCAGGAATCTCTAC 3398
Db 241 AAAATGCTTCTCTGCTGATATCAGCAGCTATAATATATTTCCACCAACTCAGGAATCTCTAC 300
QY 3399 TATGATCCCAACTCTGCCCGGTCTACTAAAAGTCTAGTAAACAACTGTAAACCACTAAT 3458
Db 301 TATGATCCCAACTCTGCCCGGTCTACTAAAAGTCTAGTAAACAACTGTAAACCACTAAT 360
QY 3459 GACAAAATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATGCGGAGAA 3518
Db 361 GACAAAATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATGCGGAGAA 420
QY 3519 GAAGAAATGCTTTTAAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGAT 3578
Db 421 GAAGAAATGCTTTTAAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGAT 480
QY 3579 GACATAAGAAAACCCCTAGGAGTTTCTTCCCTGGAATGACAACTTGCACCAATCAT 3638
Db 481 GACATAAGAAAACCCCTAGGAGTTTCTTCCCTGGAATGACAACTTGCACCAATCAT 540
QY 3639 AAAGATGCTCAGACAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATATA 3698
Db 541 AAAGATGCTCAGACAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATATA 600
QY 3699 CTTTCCCAATTTGAATCTCCAGAGAGTATCGAAACCG-TTTCTTCAATATGATGAGCT 3757
Db 601 CTTTCCCAATTTGAATCTCCAGAGAGTATCGAAACCGTTTCTTCTATATGATGAGCT 660
QY 3758 GCAGGAATGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAAC 3817
Db 661 GCAGGAATGAGGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAAC 720
QY 3818 ATTGATTATGTTTACTATATTTCTCATTTTTTGTGAGAGGTTAATTT-GCATTTAAGC-GG 3875
Db 721 ATTGATTATGTTTACTATATTTCTCATTTTTTGTGAGAGGTTAATTTGCACTTAAGCGG 780
QY 3876 AAGATATTTTCCAGAGAGGAGGATACACAAAGAG 3909
Db 781 AAGATATTTTCCAGAGAGGAGGATACCCCAAGAG 814
|||||

RESULT 9
CD657425
LOCUS
DEFINITION AGENCOURT_14538266 NIA Human H1 Embryonic Stem Cell cDNA Library
(Long) Homo sapiens cdna clone IMAGE:30419981 5', mRNA sequence.
CD657425
ACCESSION CD657425.1 GI:31899640
VERSION CD657425.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 867)
NIH-MGC <http://mgc.ncbi.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
cDNA Library Preparation: Yulan Piao and Minoru Ko
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: AGC Court Bioscience Corporation
Clone distribution: MGC C lone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM495 row: h column: 06
High quality sequence stop: 657.

FEATURES
source

1..867
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30419981"
/tissue_type="Embryonic Stem cells"
/cell_line="WA01"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"
/note="vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI;
This is a long-transcript enriched cDNA library (Genome
Res. 11: 1553-1558 (2001)). [PMID: 11544199] from WA01
cell line. Undifferentiated human ES cell line WA01/H1
was obtained from WiCell Research Institute, Inc.,
Madison, WI, cultured according to their instructions, on
MEF feeders. They formed round colonies with defined edges
and were positive for alkaline phosphatase, SSEA-4, OCT3,
OCT4, REX1, UTF, TERT, SOX2, CX43 and CX45. They are
negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1,
TUBB3, NES, GFAP, and BOMES. When confluent (18-10 days
after plating), the ES cells from 4 x 6 cm dishes were
treated with 1 mg/ml collagenase, type IV
(Invitrogen/GIBCO) for 5-10 min and gently scraped off
with 5 ml pipette. RNA was purified with TRIzol Reagent
from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558
(2001). [PMID: 11544199] Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-PGATGACTTGTAGATCCGAGCGCGCCCTTTTCTTTT-3'] from
3.4g of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lona-linker L1-Sal4, purified by phenol/chloroform
extraction, and separated from free linkers by
Centricon-100 column. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S for 25 cycles. The products
were purified by phenol/chloroform extraction and
Centricon-100 column. The cDNAs were digested with SalI
and NotI enzymes and cloned into SalI/NotI site of
pCMV-SPORT6 plasmid vector. The average insert size is
about 3.6kb."

ORIGIN

Query Match 13.4%; Score 751.8; DB 6; Length 867;
Best Local Similarity 96.7%; Pred. No. 1.8e-110;
Matches 773; Conservative 0; Mismatches 25; Indels 1; Gaps 1;
QY 2616 GTGGAACTCACCCCAAAACCATAGCGGGAATGTGACAAAGAAAGCCGCCATCT 2675
DB 11 GTGGAACTCACCCCAAAACCATAGCGGGAATGTGACAAAGAAAGCCGCCATCT 70
QY 2676 CTGATTGTTCCTGGAAAGCCAGATGACAAAGAAAGAAATCACAGGGAAGAAAA 2735

Db 71 CTGATTGTTCCTGGAAAGCCAGATGACAAAGAAAGAAATCACAGGGAAGAAAA 130
QY 2736 GAGAACAGTGAATGGAGGAAATGCTGAAATACACATAGGCGCTTACTGAAGTGTACTT 2795
Db 131 GAGAACAGTGAATGGAGGAAATGCTGAAATACACATAGGCGCTTACTGAAGTGTACTT 190
QY 2796 GGAAGAAAGCTGAGCAGATTACACAGATAGTACTTGGGCTTTTGGCATGGGAGAAAAA 2855
Db 191 GGAAGAAAGCTGAGCAGATTACACAGATAGTACTTGGGCTTTTGGCATGGGAGAAAAA 250
QY 2856 AAGTATTTCCTCAAGATCTTCTCGAGGAAGAGAGTCAATTGAAGACACAATTTGGCATACTTC 2915
Db 251 AAGTATTTCCTCAAGATCTTCTCGAGGAAGAGAGTCAATTGAAGACACAATTTGGCATACTTC 310
QY 2916 ACTGATAGCAAAATATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATAT 2975
Db 311 ACTGATAGCAAAATATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATAT 370
QY 2976 GTAATAAATTTCTAATAGCAAGTTTGGATTCACATCGCGGAAGTCCCTGCTCACATG 3035
Db 371 GTAATAAATTTCTAATAGCAAGTTTGGATTCACATCGCGGAAGTCCCTGCTCACATG 430
QY 3036 CCTCACATGATTGACCGGATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTT 3095
Db 431 CCTCACATGATTGACCGGATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTT 490
QY 3096 GACAAGACGTCATTTCAAAAGTGGCCATTTGAGGATATGAGATGAGTTCCTCTTAT 3155
Db 491 GACAAGACGTCATTTCAAAAGTGGCCATTTGAGGATATGAGATGAGTTCCTCTTAT 550
QY 3156 TTTTATTATCTCATGATGCGAGTGCAGCTGACCGCACTGAATATATCTCAAGTCTTTGATGAAGTT 3215
Db 551 TTTTATTATCTCATGATGCGAGTGCAGCTGACCGCACTGAATATATCTCAAGTCTTTGATGAAGTT 610
QY 3216 GATACAGATCAATCTGCTGCTTCTGTCAGAGAAATCCGAACACCTGGCTACAGAAAT 3275
Db 611 GATACAGATCAATCTGCTGCTTCTGTCAGAGAAATCCGAACACCTGGCTACAGAAAT-T 669
QY 3276 CACGAACCTGCGTTAAGTTTGCAGGATTCGAGGATTCGGAACACATGCTTAATAAATGTC 3335
Db 670 CACGAACCTGCGGTTAGTTTGCAGGATTCGAGGATTCGGAACACATGCTTAATAAATGTC 729
QY 3336 TCAAAATGCTTCTGCTGATATCACGAGTAAATAATATTCACCACTCAGGAATCC 3395
Db 730 CTCAAANNCTCTGCTGATATCACGAGTAAATAATATTCACCACTCAGGAATCC 789
QY 3396 TACTATGATCCCAACCTGTC 3414
Db 790 TACTATGATCCCAACCTGTC 808
RESULT 10
BQ720582
LOCUS
DEFINITION
IMAGE:6193590 5', mRNA sequence.
ACCESSION
BQ720582
VERSION
BQ720582.1 GI:21859479
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 929)
NIH-MGC <http://mgc.ncbi.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM13597 row: h column: 07
High quality sequence stop: 693.

FEATURES

Location/Qualifiers
1..929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6193590"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_libs="Lupski, sympathetic trunk"
/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
Not1; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

ORIGIN

Query Match 13.3%; Score 746.8; DB 5; Length 929;
Best Local Similarity 93.8%; Pred. No. 1.1e-109;
Matches 845; Conservative 0; Mismatches 43; Indels 13; Gaps 6;

QY 3313 TGGACACATGCTAATAAATGGCTCAAAAATGCTTCTGCTGATATCAGCAGCTAATA 3372
Db 1 TGGACACATGCTAATAAATGGCTCAAAAATGCTTCTGCTGATATCAGCAGCTAATA 60

QY 3373 ATATCCACCAACTCAGGAATCTTACTATGATATCCCAACTGCCACCGGTCACTAAAGTC 3432
Db 61 ATATCCACCAACTCAGGAATCTTACTATGATATCCCAACTGCCACCGGTCACTAAAGTC 120

QY 3433 TAGTAACAACTGTAAACCACTACTGCAAAATCCCAAGATATTAAGACAAACA 3492
Db 121 TAGTAACAACTGTAAACCACTACTGCAAAATCCCAAGATATTAAGACAAACA 180

QY 3493 AATATAGTTTGAATCATGGAGAGAAATCGCTTTTAAATGATTTCGTACCAACG 3552
Db 181 AATATAGTTTGAATCATGGAGAGAAATCGCTTTTAAATGATTTCGTACCAACG 240

QY 3553 TTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAAACCCCTAGGAAGTTTGTTCGC 3612
Db 241 TTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAAACCCCTAGGAAGTTTGTTCGC 300

QY 3613 TGAATGACAACTTGACCAATCATAAAGATGCTCAGACAGTGAAGCTGTCTCAGGG 3672
Db 301 TGAATGACAACTTGACCAATCATAAAGATGCTCAGACAGTGAAGCTGTCTCAGGG 360

QY 3673 ACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTCCCAAGAGATGCGAA 3732
Db 361 ACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTCCCAAGAGATGCGAA 420

QY 3733 ACCGTTTCCTTCATATGATGAGCTGAGGAATGGGGCTTATCCGAGACAAATTCGAGT 3792
Db 421 ACCGTTTCCTTCATATGATGAGCTGAGGAATGGGGCTTATCCGAGACAAATTCGAGT 480

QY 3793 TTGGACCCATGTGTACTAGCAACATGATTATGTTTACTATATTTCTCATTTTGTGCTG 3852
Db 481 TTGGACCCATGTGTACTAGCAACATGATTATGTTTACTATATTTCTCATTTTGTGCTG 540

QY 3853 AGCAGTTAATGCACTTAAGCGGAGATATTTCCAGAGGAGATACACAAAGAGCTA 3912
Db 541 AGCAGTTAATGCACTTAAGCGGAGATATTTCCAGAGGAGATACACAAAGAGCTA 600

QY 3913 GTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAAAACCATCTACTCAGCATTTA 3972
Db 601 GTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAAAACCATCTACTCAGCATTTA 660

QY 3973 CTGAGCATTTTAAACTCAGCTTCACAGAGATGCTTTCTGTGATGCA---TGCTTAGCAG 4029
Db 661 CTGAGTATTTTAAACTCAGCTTCACAGAGATGCTTTCTGTGATGCA---TGCTTAGCAGT 720

QY 4030 TTGGCCCGAAGAGAAATATCCAGTACC---ATGCTGTTTGTGGCATGAATATAGC 4086
Db 721 TTGGCCCGAAGAGAAATATCCAGTACCAGTACCCATGCTTTTGTGGCATGAATATAGC 780

QY 4087 CCACCTGACTA-GGAATTATTAACCAACCCACTGAAAACCTCTGT---GTCGAGCAGCTC 4142
Db 781 CCACTGACCAAGGAATTTTAAACCAACCCGCGGAAAACCTTTGTGTGGTTGAGCAGCTC 840

QY 4143 TG-AACTGATTTTACTTTTAAAGAAATTG--CTCATGGACCTGTCTCATCTTTTATAAA 4199
Db 841 TGAACACTGATTTTACTTTTAAAGAAATTGGCTCTGGGACTTGTCTATCCCTTTTATAA 900

QY 4200 A 4200
Db 901 A 901

RESULT 11
BG214021
LOCUS 783 bp mRNA linear EST 21-APR-2001
DEFINITION RST33648 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG214021
VERSION BG214021.1 GI:13735708
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 783)
AUTHORS Harrington,J.J., Sherif,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,
Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
Offenbacher,J., Danig,J. and Ducar,W.
TITLE Creation of genome-wide protein expression libraries using random
activation of genome-wide protein expression libraries using random
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL MEDLINE 21227151
PURVED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 522.
Location/Qualifiers
1..783
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_libs="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

FEATURES
source
1..783

ORIGIN
Query Match 13.3%; Score 743.8; DB 4; Length 783;
Best Local Similarity 98.7%; Pred. No. 3.5e-109;
Matches 771; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

3765 TGGGGGCTTATCGAGCAAAATTGAAGTTTGGACCATTTGTGTACTAGCACATTTGATT 3824
 Db 3 TGGGGGCTTATCGAGCAAAATTGAAGTTTGGACCATTTGTGTACTAGCACATTTGATT 61
 3825 ATGTTTACTATTTCTCAATTTTGTGAGCAGTAAATGCACTTAAGCGGAAGATATTT 3884
 Db 62 ATGTTTACTATTTCTCAATTTTGTGAGCAGTAAATGCACTTAAGCGGAAGATATTT 121
 3885 CCAGAGAGGAGATACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTC 3944
 Db 122 CCAGAGAGGAGATACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTC 181
 3945 ATTGAAACCATCTACTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGAT 4004
 Db 182 ATTGAAACCGTCTACTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGAT 241
 4005 GTCTTTGTGATGTGATGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 4064
 Db 242 GTCTTTGTGATGTGATGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 301
 4065 TGTTTCTGCGCATGAATATAGCCCACTGACTAGCAATTTTAAACCAACCCACTGAAAC 4124
 Db 302 TGTTTCTGCGCATGAATATAGCCCACTGACTAGCAATTTTAAACCAACCCACTGAAAC 361
 4125 TTGTGTCTGAGCAGCTCTGCACTGATTTTACTTTTAAAGAAATTTGCTCATGGAACCTGTC 4184
 Db 362 TTGTGTCTGAGCAGCTCTGCACTGATTTTACTTTTAAAGAAATTTGCTCATGGAACCTGTC 421
 4185 ATCTTTTAAAGAGCTCTGCACTGAGCAGCTGTTTAAATTTCCACAGCATCAT 4244
 Db 422 ATCTTTTAAAGAGCTCTGCACTGAGCAGCTGTTTAAATTTCCACAGCATCAT 481
 4245 TGCAGACTAATTTTATGAGAGAGCTCTGCACTGAGGAGTGAATGTTTAAAGAGCTCTC 4304
 Db 482 TGCAGACTAATTTTATGAGAGAGCTCTGCACTGAGGAGTGAATGTTTAAAGAGCTCTC 541
 4305 AGTCTTTCGATTCGAAAGCTTTTGTCTTAAAGTTTGGAC--TTTTTTTTTTCATTTCCC 4362
 Db 542 AGTCTTTCGATTCGAAAGCTTTTGTCTTAAAGTTTGGAC--TTTTTTTTTTCATTTCCC 601
 4363 ATTTTAAAGTGTGATTAAGTAACTAGTATTTCTTCTCTCAGTATTAACGAAATGGGA 4422
 Db 602 ATTTTAAAGTGTGATTAAGTAACTAGTATTTCTTCTCTCAGTATTAACGAAATGGGA 561
 4423 TGTCTAAACCTATTTTATAGATGTTTAAATTAATGAGCAATATACCTCTTATGA 4482
 Db 662 TGTCTAAACCTATTTTATAGATGTTTAAATTAATGAGCAATATACCTCTTATGA 721
 4483 CAATACCTAATTAATGAGTATTTTAAATTAATGAGCAATATACCTCTTAAACCACTAA 4542
 Db 722 CAATACCTAATTAATGAGTATTTTAAATTAATGAGCAATATACCTCTTAAACCACTAA 781
 4543 C 4543
 Db 782 C 782

RESULT 12
 CD652380
 LOCUS CD652380 819 bp mRNA linear EST 18-JUN-2003
 DEFINITION AGNCOURT 14537840 NIA Human H1 Embryonic Stem Cell cDNA Library
 (Long) Homo sapiens cDNA clone IMAGE:30421320 5', mRNA sequence.
 ACCESSION CD652380
 VERSION CD652380.1 GI:31889706
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 819)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

COMMENT

Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
 cDNA Library Preparation: Yulan Piao and Minoru KO
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC c lone distribution information
 can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM498 row: p column: 01
 High quality sequence stop: 629.

FEATURES
 source

1..819
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30421320"
 /tissue_type="Embryonic Stem cells"
 /cell_line="WA01"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"
 /note="vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI;
 This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]) from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on MEF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTF, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1, TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days after plating), the ES cells from 4 X 6cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID:11544199]) Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-PGACTACTTCTAGATCGAGCGCCGCTTTTCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lona-linker LL-Sal4, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN

Query Match 13.1%; Score 733.4; DB 6; Length 819;
 Best Local Similarity 94.6%; Pred. No. 1.6e-107;
 Matches 769; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
 QY 3891 AGGAGGATACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTGA 3950
 Db 1 AGAGGTATCCCAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTGA 60
 QY 3951 AAACCATCTACCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGATGCTTTT 4010
 Db 61 AAACCATCTACCTCAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGATGCTTTT 120
 QY 4011 GTGATGTGATGCTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGCTGTTTT 4070
 Db 121 GTGATGTGATGCTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGCTGTTTT 180

```
Qy 4071 GTGGCATGATATATACCCACTGCTAGCAATTTTAAACCAACCCACTGAAACTTGTGT 4130
Db 181 GTGGCATGATATATAGCCACTGACGAGAAATTTTAAACCAACCCACTGAAACTTGTGT 240
Qy 4131 GTGAGCAGCTCTGAACCTGATTTTACCTTTTAAAGAAATTTGCTCATGACCTGTCATCTTT 4190
Db 241 GTTGCAGCAGCTCTGAACCTGATTTTACCTTTTAAAGAAATTTGCTCATGACCTGTCATCTTT 300
Qy 4191 TTTTATAAAGGCTCAGTCAAGAGAGAGAGCTGTAAATTTCCCAACCAATCATTCGAGA 4250
Db 301 TTTTATAAAGGCTCAGTCAAGAGAGAGAGCTGTAAATTTCCCAACCAATCATTCGAGA 360
Qy 4251 CTAACCTTTATPAGGAGAGCTATGCCAGCTGGGAGTGATTTGCTTAAGAGGCTCCAGCTTT 4310
Db 361 CTAACCTTTATPAGGAGAGCTATGCCAGCTGGGAGTGATTTGCTTAAGAGGCTCCAGCTTT 420
Qy 4311 TGCATTCGAAAGCTTTTGTCTTAAAGTTTGGACCTTTTTCATTTCCCATTTTAA 4370
Db 421 TGCATTCGAAAGCTTTTGTCTTAAAGTTTGGACCTTTTTCATTTCCCATTTTAA 479
Qy 4371 GTAGTCTAAGTTAACTAGTATTTCTTCTCTGAGTATGAAGAAATGGGATGCTTAA 4430
Db 480 GTAGTCTAAGTTAACTAGTATTTCTTCTCTGAGTATGAAGAAATGGGATGCTTAA 539
Qy 4431 CCTATTTTATAGATGTTATTTAAATATGAGCAATATCACTCTTTATTTGACAATACCT 4490
Db 540 CCTATTTTATAGATGTTATTTAAATATGAGCAATATCACTCTTTATTTGACAATACCT 599
Qy 4491 AAATATGAGTTTATTAATATTTAAGACCTGAATGGTCTTAAACCACTAAGTCTGAA 4550
Db 600 AAATATGAGTTTATTAATATTTAAGACCTGAATGGTCTTAAACCACTAAGTCTGAA 659
Qy 4551 GAGCTCAATGATTCACATCTGAAATGCTTTGTAATTTAGTCTTACGCTTCAAGAAATGC 4610
Db 660 GAGCTCAATGATTCACATCTGAAATGCTTTGTAATTTAGTCTTACGCTTCAAGAAATGC 719
Qy 4611 TATGATTTACGTCAGCTCTAATTTTCAAGGCTAGATTTAGTCTTACCTACTTACAGATGT 4670
Db 720 TATGATTTACGTCAGCTCTAATTTTCAAGGCTAGATTTAGTCTTACCTACTTACAGATGT 779
Qy 4671 AATTATGTTTGGAAATGTCATATTTCAACAG 4703
Db 780 TGTATGTTTGGAAATGTCATATTTCAACAG 812

RESULT 13
BU933148 826 bp mRNA linear EST 18-OCT-2002
LOCUS AGENCOURT_10473204 NIH_MGC_127 Homo sapiens cDNA clone
DEFINITION IMAGE:6673871 5', mRNA sequence.
ACCESSION BU933148
VERSION BU933148.1 GI:24121967
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW2956 row: k column: 23
High quality sequence stop: 516.
```

```

Db      603 ATACTCATTTGGTATATCCAGTAGCTATCTCTCTCAGTTGGTTTTTGTATAGACAGAGAG 662
Qy      4827 GCCAGCAAC-TTTCTTTGTAAGAGCTGGTTAGTAAATATTATGAGGCCACCTG-TGTC 4884
Db      663 GCCAGCAACCTTTCTTTGTAAGAGCTGGTTAGTAAATATTATGAGGCCACCTGNTGNC 722
Qy      4885 TTGTCATACATCTCTCTG--CTGTTGTTAGTTGTTGTTTTTTTCAAAACACCTCTAA 4942
Db      723 TTGTCATACATCTCTCTGCTGCTGTTAAATTTGTTTTTTTCAAAACACCTCTAA 782
Qy      4943 AAATGTAACCAACCATGTTTA--GCTTCAGCTGTACAAAA 4981
Db      783 AAATGTAACCAACCATGTTTAAAGCTTGCAGCTGTACAAAA 823

RESULT 14
BG196325
LOCUS      773 bp mRNA linear EST 21-APR-2001
DEFINITION RST15540 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG196325
VERSION     BG196325.1 GI:13718012
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 773)
AUTHORS     Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
            Creation of genome-wide protein expression libraries using random
            activation of gene expression
            Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL     21227151
MEDLINE     11329013
PUBMED      11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 550.
            Location/Qualifiers
FEATURES    source
            1..773
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /cell_line="HT1080"
            /clone_lib="Athersys RAGE Library"
            /notes="See 'Creation of Genome-wide Protein Expression
            Libraries using Random Activation of Gene Expression',
            Nature Biotechnology, in press. Note that even though the
            cell type indicated is HT1080, since a random activation
            method was used, these sequence tags are not necessarily
            expressed in HT1080 under normal circumstances."
ORIGIN
Query Match      12.8%; Score 715.2; DB 4; Length 773;
Best Local Similarity 97.5%; Pred. No. 1.3e-104;
Matches 750; Conservative 0; Mismatches 9; Indels 10; Gaps 2;
Qy      3766 GGAGGGCTTATCGAGCAAAATTGAAGTTTGGACCATTTGTGCTAGCAACATTGATTA 3825
Db      4 GGAGGGCTTATCGAGAAAAATTGAAGTTTGGACCATTTGTGCTAGCAACATTGATTA 63
Qy      3826 TGTTCATATATCTCATTTTTTGTCTGAGC-----AGTTAATTCACCTTAAGCGGAG 3878
Db      64 TGTTCATATATCTCATTTTTTGTCTGAGCAGCATTTAGTTAATTCACCTTAAGCGGAG 123
Qy      3879 ATATTTCCAGAGGAGGATACAAAAAGAGCTAGTCCCAATCGAATCAGATATAGAG 3938

```

```

Db      124 ATATTTCCAGAGGAGGATACAAAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAG 183
Qy      3939 ATCTTCATTTGAAACCACTTACCTCAGCATTTTACTGAGCATTTTAAATTCAGCTTCAC 3998
Db      184 ATCTTCATTTGAAACCACTTACCTCAGCATTTTACTGAGCATTTTAAATTCAGCTTCAC 243
Qy      3999 AGAGATGTCCTTTGTGATGCTGATGCTTAGCAGTTTGGCCCGGAGAGAAATATCCAGTA 4058
Db      244 AGAGATGTCCTTTGTGATGCTGATGCTTAGCAGTTTGGCCCGGAGAGAAATATCCAGTA 303
Qy      4059 CCATGCTGTTTTTGTGGCATGAATATAGCCCACTGACCTAGGAAATATTATTAACCAACCCACT 4118
Db      304 CCATGCTGTTTTTGTGGCATGAATATAGCCCACTGACCTAGGAAATATTATTAACCAACCCACT 363
Qy      4119 GAAACACTTGTGTGCGAGCAGCTCTGAACCTGATTTTACTTTTTTAAAGAAATTTGCTCATGGA 4178
Db      364 GAAACACTTGTGTGTGAGCAGCTCTGAACCTGATTTTACTTTTTTAAAGAAATTTGCTCATGGA 423
Qy      4179 CCTGTCATCTCTTTTATAAAAGGCTCAGCTGACAGAGACAGCTGTTAAATTTCCACAGC 4238
Db      424 CCTGTCATCTCTTTTATAAAAGGCTCAGCTGACAGAGACAGCTGTTAAATTTCCACAGC 483
Qy      4239 AATCATTCAGACTAACTTTTATTAGGAGAGCCCTATGCCAGCTGGAGTGTCTTAAGA 4298
Db      484 AATCATTCAGACTAACTTTTATTAGGAGAGCCCTATGCCAGCTGGAGTGTCTTAAGA 543
Qy      4299 GGCTCCAGCTCTTCGATTCGAAAGCCTTTGCTAAAGTTTGGAC---TTTTTTTTTTC 4355
Db      544 GGCTCCAGCTCTTCGATTCGAAAGCCTTTGCTAAAGTTTGGACCTTTTTTTTTTTC 603
Qy      4356 ATTTCCTCATTTTAAAGTAGTTACTAAAGTAACTAGTTATTCTTCTCTCTGAGTATAACGA 4415
Db      604 ATTTCCTCATTTTAAAGTAGTTACTAAAGTAACTAGTTATTCTTCTCTCTGAGTATAACGA 663
Qy      4416 ATTGGATGCTAAACCTTTTATTAGATGTTATTTAAATATGACGACATATCACCTC 4475
Db      664 ATGGGATGCTAAACCTTTTATTAGATGTTATTTAAATATGACGACATATCACCTC 723
Qy      4476 TTATTGACATACCTAAATATGATGTTTTTATTATTTAAAGACTGTA 4524
Db      724 TTATTGACATACCTAAATATGATGTTTTTATTATTTAAAGACTGTA 772

RESULT 15
BG194321
LOCUS      795 bp mRNA linear EST 21-APR-2001
DEFINITION RST13466 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG194321
VERSION     BG194321.1 GI:13716008
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 795)
AUTHORS     Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
            Creation of genome-wide protein expression libraries using random
            activation of gene expression
            Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL     21227151
MEDLINE     11329013
PUBMED      11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 441.

```


FEATURES
source

Location/Qualifiers

1..795
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

ORIGIN

Query Match 12.7%; Score 710.8; DB 4; Length 795;
Best Local Similarity 97.1%; Pred. No. 6.8e-104;
Matches 775; Conservative 0; Mismatches 16; Indels 7; Gaps 5;

QY 3765 TGGAGGCTTATCGAGCAAAATTGAAGTTTTGGACCCATTGTGTACTAGCAACATTGATT 3824
Db 4 TGGAGGACTATCGAGCAAAATTG-AGTTTTGGACCCATTGTGTACTAGCAACATTGATT 62

QY 3825 ATGTTTACTATATTCTCATTTTTTGTCTGAGCAGTTAATTGCACCTTAAGCGGAAGATATTT 3884
Db 63 ATGTTTACTATATTCTCATTTTTTGTCTGAGCAGTTAATTGCACCTTAAGCGGAAGATATTT 122

QY 3885 CCCAGAGGAGGATACAAAGAAGCTAGTCCCAATCGATCAGATATAGAAGATCTTC 3944
Db 123 CCCTGAAGAGGATACAAAGAAGCTAGTCCCAATCGATCAGATATAGAAGATCTTC 182

QY 3945 ATTTGAAACCATTCTACCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCACAGAGAT 4004
Db 183 ATTTGAAACCATTCTACCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCACAGAGAT 242

QY 4005 GTCCTTTGTGATGTCATGCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGC 4064
Db 243 GTCCTTTGTGATGTCATGCTTAGCAGTTTGGCCCGAAGAGAAATAT-CAGTACCATGC 301

QY 4065 TGTTTTGTGGCATGAATATAGCCACCTGACTAGGAATTTTACCACCCACTGGAAC 4124
Db 302 TGTTTTGTGGCATGAATATAGCCACCTGACTAGGAATTTTACCACCCACTGGAAC 361

QY 4125 TTGTGTGCGAGCAGCTCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGTC 4184
Db 362 TTGTGTGTTGAGCAGCTCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGTC 421

QY 4185 ATCCTTTTATAAAGCTCACTGACAGAGACAGCTGTTAATTTCCACAGCAATCAT 4244
Db 422 ATCCTTTTATAAAGCTCACTGACAGAGACAGCTGTTAATTTCCCAACAACATCAT 481

QY 4245 TGCAGACTAATTTTATTAGGAGAGCCCTATGCCAGCTGGGAGTGATTGCTAAGAGGCTCC 4304
Db 482 TGCAGACTAATTTTATTAGGAGAGCCCTATGCCAGCTGGGAGTGATTGCTAAGAGGCTCC 541

QY 4305 AGCTTTTCCTCCAAAGCCCTTTGCTAAAGTTTGGACCTTTTTTTTTTTCATTTCCCAT 4364
Db 542 AGCTTTTCCTCCAAAGCCCTTTGCTAAAGTTTGGACCTTTTTCATTTCCCAT 598

QY 4365 TTTTAAAGTAGTACTAAGCTAAGTATTCTTCTGCTTCTGAGTATAACGAATTTGGGATG 4424
Db 599 TTTTAAAGTAGTACTAAGCTAAGTATTCTTCTGCTTCTGAGTATAACGAATTTGGGATG 658

QY 4425 TCTAAACCTATTTTATAGATGTTATTTAAATAATGACGAATATACCTCTTATTGACA 4484
Db 659 TCTAAACCTA-TTTTATAGATGTTATTTAAATAATGACGAATATACCTCTTATTGANC 717

QY 4485 ATACCTAAATATGAGTTTTTATTAATTTAAGACTGTAATGCTTTAAACCACTAACT 4544
Db 718 ATACCTAAATATGAGTTTTTATTAATTTAAGACTGTAATGCTTTAAACCACTAACT 776

QY 4545 ACTGAAGAGCTCAATGAT 4562
|||||

Db 777 ACTGAAGAGCTCAATGAT 794

Search completed: November 22, 2004, 10:37:19
Job time : 15717 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:38:18 ; Search time 106.597 Seconds
(without alignments)
3122.980 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907
Sequence: 1 MLFKLQRQYTCLSHRYGL.....SLKTQLAYFTDKNTRGQLX 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A Geneseq_23Sep04:*

- 1: geneseq1980s:*
- 2: geneseq1990s:*
- 3: geneseq2000s:*
- 4: geneseq2001s:*
- 5: geneseq2002s:*
- 6: geneseq2003as:*
- 7: geneseq2003bs:*
- 8: geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4307	100.0	928	4	ABR61376 Human Glc
2	4307	100.0	928	7	ADD27812 Human Glc
3	4307	100.0	928	7	ABW01488 Human Glc
4	4307	100.0	928	7	ABW01537 Human Glc
5	4896	99.8	1256	5	Aae25290 Human nuc
6	4300	94.4	1199	7	ADD27810 Soluble h
7	4300	94.4	1199	7	ABW01487 N-acetyl
8	4630	94.4	1199	7	ABW01536 N-acetyl
9	4500	91.7	1196	5	Aae25294 Human nuc
10	4468	91.1	846	6	ADA54824 Human pro
11	4461	90.9	847	7	ADG39856 Protein s
12	4391	89.5	1459	6	ABU07381 Human pro
13	4391	89.5	1459	7	ADG39794 Human nov
14	3894	79.4	908	7	ABR61384 Murine gl
15	3894	79.4	908	7	ADD27817 GLcNAc-ph
16	3894	79.4	908	7	ABW01491 Mouse pro
17	3894	79.4	908	7	ABW01540 Mouse pro
18	2016	41.1	663	7	ADG39855 Protein s
19	1900	38.7	367	4	AM16457 Peptide #
20	1900	38.7	367	4	AAW28952 Peptide #
21	1900	38.7	367	4	ABE30279 Peptide #
22	1900	38.7	367	4	AAW56268 Human bra
23	1900	38.7	367	4	AAW04186 Peptide #
24	1900	38.7	367	5	ABG38224 Human pep
25	1900	38.7	367	8	ABO59293 Human gen

ALIGNMENTS

RESULT 1

ABR61376
ID ABR61376 standard; protein; 928 AA.

XX ABR61376;

XX AC

XX 01-AUG-2003 (first entry)

XX Human GlcNAc-phosphotransferase alpha-subunit.

XX Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;

KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;

KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;

KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;

KW lysosomal storage disease; enzyme; alpha-subunit.

XX Homo sapiens.

XX US6537785-B1.

XX 25-MAR-2003.

XX 10-AUG-2000; 2000US-00635077.

XX 14-SEP-1999; 99US-0153831P.

XX (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX Canfield WM;

XX WFI; 2001-290356/30.

XX N-PSDB; ACC81001.

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX Disclosure; Page 26-28; 62pp; English.

XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester

XX alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-

XX Acetylglucosaminidase) (II). The protein of the invention has

XX nephrotropic activity, and may be useful in enzyme replacement therapy. A

XX protein of the invention (I), (II) is useful for preparing a

XX phosphorylated lysosomal hydrolase. The phosphorylated hydrolase

XX comprising a terminal mannose-6-phosphate, is useful for treating a

Add26807 Human adi
Abu11844 Human MD
Adg39857 Protein s
Abri61382 Partial r
Add27822 Rat alpha
Abw01494 Partial r
Abw01543 Partial r
Adg39859 Protein s
AB59094 Drosophila
ABri61383 Partial D
Add27824 Fruit fly
Abw01495 Partial f
Abw01544 Partial f
ABO59064 Human gen
Adi21059 Novel num
AB338187 Peptide #
Aam31616 Peptide #
Aam71337 Human bon
Aam58821 Human bra
ABG53045 Human liv

CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention

XX SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 4; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFKLLQRTQVTCLSHRYGLVYVCFGLGVVTVISAFQGEVLEWSDQYHVLFDSDYRNI 60
 Db 1 MLFKLLQRTQVTCLSHRYGLVYVCFGLGVVTVISAFQGEVLEWSDQYHVLFDSDYRNI 60
 Qy 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELLEKQVREOQMEEEKAMREILGKNTTEPTK 120
 Db 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELLEKQVREOQMEEEKAMREILGKNTTEPTK 120
 Qy 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180
 Db 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180
 Qy 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMDLAFSLGPPPTFKET 240
 Db 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMDLAFSLGPPPTFKET 240
 Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA 300
 Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA 300
 Qy 301 YLLWDLAISQSKODEDISASRPEDNEELYSIRSHAPWVRNFIPTNGQIPSWLNL 360
 Db 301 YLLWDLAISQSKODEDISASRPEDNEELYSIRSHAPWVRNFIPTNGQIPSWLNL 360
 Qy 361 DNPRTVITVHQVDPNLSHPTFSSPAIESHRIEGLSKFIYLANDVDMFGKVPDDF 420
 Db 361 DNPRTVITVHQVDPNLSHPTFSSPAIESHRIEGLSKFIYLANDVDMFGKVPDDF 420
 Qy 421 YSHSGKQVLTVPNCAECPGSIWKDGYCDKACNNSACDWDGDCSNGSGSYIAG 480
 Db 421 YSHSGKQVLTVPNCAECPGSIWKDGYCDKACNNSACDWDGDCSNGSGSYIAG 480
 Qy 481 GGGTSGIVGHPWQFGGINSVSCNCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540
 Db 481 GGGTSGIVGHPWQFGGINSVSCNCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540
 Qy 541 HELYKVILLNQTHYIIPKGECLPYFSPAFAVAKRGEVGYSDNPIIRHASIANKWTIHL 600
 Db 541 HELYKVILLNQTHYIIPKGECLPYFSPAFAVAKRGEVGYSDNPIIRHASIANKWTIHL 600
 Qy 601 IMHSGMNATTHENLTFTQNTNDEEFKQIIVEDVTRGPKLNSTAGKYNLSPITLLP 660
 Db 601 IMHSGMNATTHENLTFTQNTNDEEFKQIIVEDVTRGPKLNSTAGKYNLSPITLLP 660
 Qy 661 EAEILFEDIPEKKEFPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQSLNLTDLQLEH 720
 Db 661 EAEILFEDIPEKKEFPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQSLNLTDLQLEH 720
 Qy 721 GDITLKGNLSKALLSFILNNSOAKIKNOAIIITDETNDLSVAPQKQVHKSILPNSLG 780
 Db 721 GDITLKGNLSKALLSFILNNSOAKIKNOAIIITDETNDLSVAPQKQVHKSILPNSLG 780
 Qy 781 VSERLQRLTTPAVSVKNGHQDQGNPPLDLETTARFRVETHQTKTIGGNVTEKPPSLIV 840
 Db 781 VSERLQRLTTPAVSVKNGHQDQGNPPLDLETTARFRVETHQTKTIGGNVTEKPPSLIV 840
 Qy 841 PLESQMTKEKKITCKEKENSRMEENAHNGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 Db 841 PLESQMTKEKKITCKEKENSRMEENAHNGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 Qy 901 QDLIDDEESLKTQLAYFTDSKNTGRQJK 928
 Db 901 QDLIDDEESLKTQLAYFTDSKNTGRQJK 928

RESULT 2

ADD27812
 ID ADD27812 standard; protein; 928 AA.

XX AC ADD27812;
 XX AC ADD27812;
 DT 15-JAN-2004 (first entry)
 XX Human GlcNAc-phosphotransferase alpha subunit.
 DE human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
 KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.
 XX Homo sapiens.
 OS Homo sapiens.
 PN US2003119088-A1.
 XX 26-JUN-2003.
 XX 21-DEC-2001; 2001US-00023888.
 XX 21-DEC-2001; 2001US-00023888.
 XX (NOVA-) NOVATIME PHARM INC.
 PA Canfield W, Kudo M;
 PI WPI; 2003-801323/75.
 DR N-PSDB; ADD27811.
 XX

PT Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.
 XX

PS Claim 7; SEQ ID NO 4; 55pp; English.
 XX

CC The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of the human GlcNAc-phosphotransferase alpha subunit.
 XX

SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFKLLQRTQVTCLSHRYGLVYVCFGLGVVTVISAFQGEVLEWSDQYHVLFDSDYRNI 60
 Db 1 MLFKLLQRTQVTCLSHRYGLVYVCFGLGVVTVISAFQGEVLEWSDQYHVLFDSDYRNI 60
 Qy 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELLEKQVREOQMEEEKAMREILGKNTTEPTK 120
 Db 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELLEKQVREOQMEEEKAMREILGKNTTEPTK 120
 Qy 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180
 Db 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180
 Qy 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMDLAFSLGPPPTFKET 240
 Db 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMDLAFSLGPPPTFKET 240
 Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA 300
 Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA 300
 Qy 301 YLLWDLAISQSKODEDISASRPEDNEELYSIRSHAPWVRNFIPTNGQIPSWLNL 360

Db 301 YLLWDLAISQSKQDEDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRTVITVTHQDVFNRLSHLPTSSPAIESHIEGLSOKFYIYNDVDFGKDVWPDFF 420
Db 361 DNPRTVITVTHQDVFNRLSHLPTSSPAIESHIEGLSOKFYIYNDVDFGKDVWPDFF 420
Qy 421 YSHSGQKQVLTWVPNCAEGCPGSKWIKDGYCDKACNNSACDMDGGDCSGNSGSRVYAG 480
Db 421 YSHSGQKQVLTWVPNCAEGCPGSKWIKDGYCDKACNNSACDMDGGDCSGNSGSRVYAG 480
Qy 481 GGGTSGISGVHPWQFGGINSVSYNOCGANSWLADFCQACNVLSGFDAGCGQDHF 540
Db 481 GGGTSGISGVHPWQFGGINSVSYNOCGANSWLADFCQACNVLSGFDAGCGQDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
Qy 721 GDITLKGYNLSKALLSFLMSQAKIKNOAIITDETNSLVAPOKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLSFLMSQAKIKNOAIITDETNSLVAPOKQVHKSILPNSLG 780
Qy 781 VSELRORLTFAVSVKNGHDQGNPLDLETTARFVETHQKTIGCNVTKEKPPSLIV 840
Db 781 VSELRORLTFAVSVKNGHDQGNPLDLETTARFVETHQKTIGCNVTKEKPPSLIV 840
Qy 841 PLESQMTKEKITGKEKENSMEENAEHIGVTEVLLGRKLOHYTDSYLGFLPWKKKYF 900
Db 841 PLESQMTKEKITGKEKENSMEENAEHIGVTEVLLGRKLOHYTDSYLGFLPWKKKYF 900
Qy 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928
Db 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 3
ABW01488
ID ABW01488 standard; protein; 928 AA.
AC ABW01488;
DT 15-JAN-2004 (first entry)
DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.
XX
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KW gastrointestinal; human; enzyme; lectin resistant cell;
KW deoxymannojirimycin; kifunensine; glycosylation inhibition.
XX
OS Homo sapiens.
XX
FN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield WM;
XX
DR WPI; 2003-810984/76.

DR N-PSDB; AAD62491.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 10; Page 15-18; 46pp; English.
XX
XX The invention relates to a method for producing a high mannose
XX glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is human N-
CC acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha subunit precursor
CC protein
XX
SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLFKLQRTYTCLSHRYGLYVCFGLGVVTVISAFQFGEVLEWRSRDQHHVLFDSYRDI 60
Db 1 MLFKLQRTYTCLSHRYGLYVCFGLGVVTVISAFQFGEVLEWRSRDQHHVLFDSYRDI 60
Qy 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREOMEEQKAMREILGNVTEPTK 120
Db 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREOMEEQKAMREILGNVTEPTK 120
Qy 121 KSEKQLECLLTHCIVKPMVLVLPALPANI TLKDVPSLYPSFHSASDIFNVAKPSTNV 180
Db 121 KSEKQLECLLTHCIVKPMVLVLPALPANI TLKDVPSLYPSFHSASDIFNVAKPSTNV 180
Qy 181 SVVVFSTKDVDEDAHSGLLKNSRQTWVRGYLTTRKEVGLVLMQDLAFLSGFPPTFKET 240
Db 181 SVVVFSTKDVDEDAHSGLLKNSRQTWVRGYLTTRKEVGLVLMQDLAFLSGFPPTFKET 240
Qy 241 NQKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Qy 301 YLLWDLAISQSKQDEDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLAISQSKQDEDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRTVITVTHQDVFNRLSHLPTSSPAIESHIEGLSOKFYIYNDVDFGKDVWPDFF 420
Db 361 DNPRTVITVTHQDVFNRLSHLPTSSPAIESHIEGLSOKFYIYNDVDFGKDVWPDFF 420
Qy 421 YSHSGQKQVLTWVPNCAEGCPGSKWIKDGYCDKACNNSACDMDGGDCSGNSGSRVYAG 480
Db 421 YSHSGQKQVLTWVPNCAEGCPGSKWIKDGYCDKACNNSACDMDGGDCSGNSGSRVYAG 480
Qy 481 GGGTSGISGVHPWQFGGINSVSYNOCGANSWLADFCQACNVLSGFDAGCGQDHF 540
Db 481 GGGTSGISGVHPWQFGGINSVSYNOCGANSWLADFCQACNVLSGFDAGCGQDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

Db 661 EAEILFEDIPEKFPKPPKRDHVNSTRRAQAEVKIPLVNISSLPAQKQSLNTLDLQLEH 720
 QY 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIIITDNDLSLVAPOKQVHKSTILPNSLG 780
 Db 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIIITDNDLSLVAPOKQVHKSTILPNSLG 780
 QY 781 VSELRQLRTPPAVSVKVNQHDQGNPPDLDTTARFRVETHQKTIIGNVTKEKPPSLIV 840
 Db 781 VSELRQLRTPPAVSVKVNQHDQGNPPDLDTTARFRVETHQKTIIGNVTKEKPPSLIV 840
 QY 841 PLESQMTKEKTIQKXKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900
 Db 841 PLESQMTKEKTIQKXKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900
 QY 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928
 Db 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 4

ABW01537 standard; protein; 928 AA.

ID ABW01537 standard; protein; 928 AA.

XX AC ABW01537;

XX DT 15-JAN-2004 (first entry)

XX DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.

XX KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;

XX KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.

XX OS Homo sapiens.

XX FN US2003124653-A1.

XX PD 03-JUL-2003.

XX PF 21-DEC-2001; 2001US-00023890.

XX PR 21-DEC-2001; 2001US-00023890.

XX PA (NOVA-) NOVAZYME PHARM INC.

XX PI Canfield WM;

XX DR WPI; 2003-810985/76.

XX DR N-PSDB; AAD62650.

XX PT Producing a glycoprotein with reduced complex carbohydrates by culturing

XX PT the lectin resistant mammalian cell expressing the glycoprotein for

XX PT treating lysosomal storage disease.

XX PS Claim 10; Page 16-18; 46pp; English.

XX CC The present invention provides a method of producing a glycoprotein

XX CC having reduced complex carbohydrates by culturing the lectin resistant

XX CC mammalian cell expressing the glycoprotein. The method is useful for

XX CC producing a glycoprotein with reduced complex carbohydrates for treating

XX CC lysosomal storage disease. The present invention is also useful in gene

XX CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-

XX CC phosphotransferase alpha subunit precursor protein

XX SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRQYTCLSHRYGLYVCFGLGVVTVISAFQGEVLEWSDQVHVLFDSDYRNI 60

Db 1 MLFKLQRQYTCLSHRYGLYVCFGLGVVTVISAFQGEVLEWSDQVHVLFDSDYRNI 60

QY 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELLKELQVREQMEEEQKAMEIILGKNTTEPTK 120
 Db 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELLKELQVREQMEEEQKAMEIILGKNTTEPTK 120
 QY 121 KSEKQLECLLTHICIKVPMVLDPALPANITLKDVPSLYPSFHSASDIPNVAKPKNPSTNV 180
 Db 121 KSEKQLECLLTHICIKVPMVLDPALPANITLKDVPSLYPSFHSASDIPNVAKPKNPSTNV 180
 QY 181 SVVVFDSKQVDEDAHSGLLKGNRSOTVMRGYLTITDKVEPGLVLMQDLAFSLGFPPTPKET 240
 Db 181 SVVVFDSKQVDEDAHSGLLKGNRSOTVMRGYLTITDKVEPGLVLMQDLAFSLGFPPTPKET 240
 QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOELNKQTKKMTIDGKELTISPA 300
 Db 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOELNKQTKKMTIDGKELTISPA 300
 QY 301 YLLWDLAISQSKODEDJSASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
 Db 301 YLLWDLAISQSKODEDJSASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
 QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDWPPDF 420
 Db 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDWPPDF 420
 QY 421 YSHSGQKQVYLTWPVPCAECPGSGWIKDGYCDKACNNSACDWDGDCSGNSGSRITAG 480
 Db 421 YSHSGQKQVYLTWPVPCAECPGSGWIKDGYCDKACNNSACDWDGDCSGNSGSRITAG 480
 QY 481 CGGTGSGIVGHVPWFQGGGINSVSYCNOGCANSLADKFCDOACNVLSGCGFDAGCGGDHF 540
 Db 481 CGGTGSGIVGHVPWFQGGGINSVSYCNOGCANSLADKFCDOACNVLSGCGFDAGCGGDHF 540
 QY 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVGAYSNDNPIIRHASIANKWKTIHL 600
 Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVGAYSNDNPIIRHASIANKWKTIHL 600
 QY 601 IMHSGMATTIHFNLTFTONTNDEBEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
 Db 601 IMHSGMATTIHFNLTFTONTNDEBEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
 QY 661 EAEILFEDIPEKFPKPPKRDHVNSTRRAQAEVKIPLVNISSLPAQKQSLNTLDLQLEH 720
 Db 661 EAEILFEDIPEKFPKPPKRDHVNSTRRAQAEVKIPLVNISSLPAQKQSLNTLDLQLEH 720
 QY 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIIITDNDLSLVAPOKQVHKSTILPNSLG 780
 Db 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIIITDNDLSLVAPOKQVHKSTILPNSLG 780
 QY 781 VSELRQLRTPPAVSVKVNQHDQGNPPDLDTTARFRVETHQKTIIGNVTKEKPPSLIV 840
 Db 781 VSELRQLRTPPAVSVKVNQHDQGNPPDLDTTARFRVETHQKTIIGNVTKEKPPSLIV 840
 QY 841 PLESQMTKEKTIQKXKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900
 Db 841 PLESQMTKEKTIQKXKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900
 QY 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928
 Db 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 5

AAE25290

ID AAE25290 standard; protein; 1256 AA.

XX AC AAE25290;

XX DT 30-OCT-2002 (first entry)

XX DE Human nucleic acid-associated protein (NAAP-9).

XX KW Human; nucleic acid-associated protein; NAAP-9; neurological disorder;

XX KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

KW

KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
KW gene therapy; nontropic; neuroprotective; cerebroprotective; virucide;
KW immunosuppressive; protozoicide; antimicrobial.

OS Homo sapiens.
FX Key Location/Qualifiers
FH Domain 448..469 /note= "Notch domain"
FT Domain 500..536 /note= "Notch domain"
FT Domain 1018..1030 /note= "EF-hand calcium-binding domain"

XX WO200250279-A2.
XX 27-JUN-2002.
XX 19-DEC-2001; 2001WO-US050256.
XX 21-DEC-2000; 2000US-0257714P.
XX 05-JAN-2001; 2001US-0260081P.
XX 16-JAN-2001; 2001US-0262302P.
XX 23-JAN-2001; 2001US-0263823P.
XX 02-FEB-2001; 2001US-0266089P.
XX 29-OCT-2001; 2001US-0348442P.
XX (INCV-) INCYTE GENOMICS INC.

XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
PI Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
PI Borowsky M, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
PI Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;
PI Swarnakar A, Reddy R;
XX WPI; 2002-519887/55.
DR N-PsDB; RAD41199.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,
PT treating and preventing cell proliferative (e.g. cancers), neurological
PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX Claim 64; Page 163-165; 193pp; English.

XX The invention relates to nucleic acid-associated proteins (NAAP) and
CC nucleic acids. The nucleic acid and amino acid sequences are useful for
CC diagnosing, treating and preventing cell proliferative e.g.
CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-9

XX Sequence 1256 AA;
Query Match 99.8%; Score 4896; DB 5; Length 1256;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 926; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLQRTVTCLSHRYGLVCFGLGVVTVTSAPQGEVLEWSRDQYHVLFDSDNI 60
DB 1 MLFKLQRTVTCLSHRYGLVCFGLGVVTVTSAPQGEVLEWSRDQYHVLFDSDNI 60
QY 61 AGKSFQNRCLCLPMPIDVVYTWNGTDLLELKLQVREQMEEEQKAMREILGKNTTPTK 120
DB 61 AGKSFQNRCLCLPMPIDVVYTWNGTDLLELKLQVREQMEEEQKAMREILGKNTTPTK 120
QY 121 KSEKQLECLLTCHIKVPMVLDPALPANITLKDPSLYPSFHSASDIFNVAKPKNPTNV 180

DB 121 KSEKQLECLLTCHIKVPMVLDPALPANITLKDPSLYPSFHSASDIFNVAKPKNPTNV 180
QY 181 SVTVFSTXDVEDAHSGLLKGNRSQTVWRYGTLTKEVPLVLMODLAFSLGPPFPFKET 240
DB 181 SVTVFSTXDVEDAHSGLLKGNRSQTVWRYGTLTKEVPLVLMODLAFSLGPPFPFKET 240
QY 241 NQLTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
DB 241 NQLTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
QY 301 YLLWDLISAISQKQDEDI SASRPEDNEELRYSLRSTIERHAPVWRNIFIVTNGQISWNL 360
DB 301 YLLWDLISAISQKQDEDI SASRPEDNEELRYSLRSTIERHAPVWRNIFIVTNGQISWNL 360
QY 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHRIHRIEGLSQFIYLNDDVMFGKVPDPDF 420
DB 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHRIHRIEGLSQFIYLNDDVMFGKVPDPDF 420
QY 421 YSHSKGQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSYIAG 480
DB 421 YSHSKGQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSYIAG 480
QY 481 GGGTSGIGVGHPMQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTSGIGVGHPMQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWTIHL 600
QY 601 IMHSGNNTIHNLTQNTNDEEFKQITVEVDTRREGPKLNSTAGKGYENLVSPITLLP 660
DB 601 IMHSGNNTIHNLTQNTNDEEFKQITVEVDTRREGPKLNSTAGKGYENLVSPITLLP 660
QY 661 EABILFEDIPKFKRFPKRDVNSTRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
DB 661 EABILFEDIPKFKRFPKRDVNSTRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
QY 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIIITDETNDLSLVAPOEKVHKSILPNSLG 780
DB 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIIITDETNDLSLVAPOEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAPVSVKVGHDQGNPDLDTTAFRVETHTKTIGGNVTKKPPSLIV 840
DB 781 VSERLQRLTFPAPVSVKVGHDQGNPDLDTTAFRVETHTKTIGGNVTKKPPSLIV 840
QY 841 PLESQMTKEKKTITGKEKENSMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
DB 841 PLESQMTKEKKTITGKEKENSMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLLDEESLTKQLAYFTDSKNTGROLK 928
DB 901 QDLLDEESLTKQLAYFTDSKNTGROLK 928

RESULT 6
ADD27810
ID ADD27810 standard; protein; 1199 AA.
XX
AC ADD27810;
XX
DT 15-JAN-2004 (first entry)
XX
XX Soluble human GlcNAc-phosphotransferase.

DE human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
KW enzyme.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX US2003119088-A1.
FN

XX 26-JUN-2003.
 XX 21-DEC-2001; 2001US-00023888.
 XX 21-DEC-2001; 2001US-00023888.
 XX (NOVA-) NOVAZYME PHARM INC.
 XX Canfield W, Kude M;
 XX WPI; 2003-801323/75.
 XX N-PSDB; ADD27809.
 XX
 XX Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.
 XX
 XX Claim 3; SEQ ID NO 2; 55pp; English.
 XX
 XX The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of soluble human GlcNAc-phosphotransferase.
 XX
 XX Query Match 94.4%; Score 4630; DB 7; Length 1199;
 XX Best Local Similarity 99.3%; Pred. No. 0;
 XX Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 45 SRDQVHVFDSYRNIAKSPQNLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEQ 104
 DB 35 SRDQVHVFDSYRNIAKSPQNLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEQ 94
 QY 105 KAMEIILGKNTTEPTKSEKOLECLTHCIVKPMVLDPALPANITLKDVPSPYPSFHS 164
 DB 95 KAMEIILGKNTTEPTKSEKOLECLTHCIVKPMVLDPALPANITLKDVPSPYPSFHS 154
 QY 165 SDIFNVAKPNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQTVMRGYLTDDKEVPGVLVM 224
 DB 155 SDIFNVAKPNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQTVMRGYLTDDKEVPGVLVM 214
 QY 225 QDLAFLSGFPPTFKETVQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOT 284
 DB 215 QDLAFLSGFPPTFKETVQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOT 274
 QY 285 KKNMTIDGKELTISPAYLLWLSAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVR 344
 DB 275 KKNMTIDGKELTISPAYLLWLSAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVR 334
 QY 345 NIFVTNGQIPSWNLNPNRVTIVTHQDVFRNLNLSHLPFTSPSPAESIHRIEGLSOKFIY 404
 DB 335 NIFVTNGQIPSWNLNPNRVTIVTHQDVFRNLNLSHLPFTSPSPAESIHRIEGLSOKFIY 394
 QY 405 LNDVFMFGKDVWDDFYSHSGQKVLITWVPNCAECGPGSWIKDGYCDKACNNSACDWD 464
 DB 395 LNDVFMFGKDVWDDFYSHSGQKVLITWVPNCAECGPGSWIKDGYCDKACNNSACDWD 454
 QY 465 GGDGSGNSGGGRIYAGGGGTGSIQVGHFWPFGGINSVSYCNOGCSANSLWADRFCDQACN 524
 DB 455 GGDGSGNSGGGRIYAGGGGTGSIQVGHFWPFGGINSVSYCNOGCSANSLWADRFCDQACN 514
 QY 525 VLSGCFDAGCGQDHFHELYKVIILLPNOTHVIIPKGECLPYFSPFAEVAKRGVSGAYSDNP 584
 DB 515 VLSGCFDAGCGQDHFHELYKVIILLPNOTHVIIPKGECLPYFSPFAEVAKRGVSGAYSDNP 574
 QY 585 IIRHASIANKWKTIIHLIMHSGMNATTIHFNLTFONTNDEEFKMQITVEVDTRGPKLNST 644
 DB 575 IIRHASIANKWKTIIHLIMHSGMNATTIHFNLTFONTNDEEFKMQITVEVDTRGPKLNST 634

QY 645 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFPKFRHDVNSTRAAQBEVKIPLVNIISLLP 704
 DB 635 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFPKFRHDVNSTRAAQBEVKIPLVNIISLLP 694
 QY 705 KDAQSLNLTLDLQLEHGDITLKGYNLSKALSRLSPLMNSQHAKIKNOAILTDETNDSLVA 764
 DB 695 KDAQSLNLTLDLQLEHGDITLKGYNLSKALSRLSPLMNSQHAKIKNOAILTDETNDSLVA 754
 QY 765 PQEKQVHKSIILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLDETARFRVETHTQK 824
 DB 755 PQEKQVHKSIILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLDETARFRVETHTQK 814
 QY 825 TIGGNVTKEKPSLIIVPLESQMTKEKKTGKEKNSRMEENAEHIGVTEVLLGRKLOHY 884
 DB 815 TIGGNVTKEKPSLIIVPLESQMTKEKKTGKEKNSRMEENAEHIGVTEVLLGRKLOHY 874
 QY 885 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNVTR 925
 DB 875 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNVTR 915
 XX
 XX RESULT 7
 XX ABW01487
 XX ID ABW01487 standard; protein; 1199 AA.
 XX AC ABW01487;
 XX DT 15-JAN-2004 (first entry)
 XX DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase.
 XX KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
 KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
 KW gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;
 KW kifunensine; glycosylation inhibition.
 XX OS Unidentified.
 XX US2003124652-A1.
 XX PN 03-JUL-2003.
 XX PD 21-DEC-2001; 2001US-00023889.
 XX PF 21-DEC-2001; 2001US-00023889.
 XX PR (NOVA-) NOVAZYME PHARM INC.
 XX PA Canfield WM;
 XX PI WPI; 2003-810984/76.
 XX DR N-PSDB; ADD62490.
 XX
 XX Producing a high mannose glycoprotein for treating lysosomal storage
 PT disease, comprises culturing the lectin resistant mammalian cell in the
 PT presence of deoxymannojirimycin and kifunensine.
 XX
 XX Claim 8; Page 10-13; 46pp; English.
 XX
 XX The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
 CC (GlcNAc)-phosphotransferase
 XX
 XX Sequence 1199 AA;
 XX SQ

Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQVHVLDSYRDNIAGSKFQNRCLPMPIDVVTWNGTDLELLKELQOVRQMEBEQ 104
DB 35 SRDQVHVLDSYRDNIAGSKFQNRCLPMPIDVVTWNGTDLELLKELQOVRQMEBEQ 94

QY 105 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLDPALPANITLKDVPSPSPHSA 164
DB 95 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLDPALPANITLKDVPSPSPHSA 154

QY 165 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLM 224
DB 155 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLM 214

QY 225 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKQT 284
DB 215 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKQT 274

QY 285 KKNMTIDGKELTISPAYLLWDLISAISQSKQDEDIASRFEDNEELRYSLRSTIERHAPVR 344
DB 275 KKNMTIDGKELTISPAYLLWDLISAISQSKQDEDIASRFEDNEELRYSLRSTIERHAPVR 334

QY 345 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIY 404
DB 335 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIY 394

QY 405 LNDDVMFGKDWPDFFYSHSGKQKYLTPVNPNCAGCGPGSWIKDGYCDKACNNSACDWD 464
DB 395 LNDDVMFGKDWPDFFYSHSGKQKYLTPVNPNCAGCGPGSWIKDGYCDKACNNSACDWD 454

QY 465 GGDCSGNSGGSGRYIAGGGGTGSGIGVGPQWQGGGINSVSYCNQGCANSWLADKFCDOACN 514
DB 525 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGVEGAYSDNP 584
DB 515 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGVEGAYSDNP 574

QY 585 IIRHASIANKWKTHLIMHSGNATTHFNLTFTQNTDDEEFKMOITVEVDTRGPKLNST 644
DB 575 IIRHASIANKWKTHLIMHSGNATTHFNLTFTQNTDDEEFKMOITVEVDTRGPKLNST 634

QY 645 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRRRAQBEVKIPLVNIISLLP 704
DB 635 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRRRAQBEVKIPLVNIISLLP 694

QY 705 KDAQLSLNTLQLEHGDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSVA 764
DB 695 KDAQLSLNTLQLEHGDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSVA 754

QY 765 PQEKQVHKSILPNSILGVSERLQRLTFPAVSVKVNHGQGNPPLDLETTARFRVETHQK 824
DB 755 PQEKQVHKSILPNSILGVSERLQRLTFPAVSVKVNHGQGNPPLDLETTARFRVETHQK 814

QY 825 TIGGNVTKRPPSLIIVLESOMTEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHY 884
DB 815 TIGGNVTKRPPSLIIVLESOMTEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHY 874

QY 885 TDSYLGFLPWEKKYKFODLDEESLKTQLAYFTDSKNTGR 925
DB 875 TDSYLGFLPWEKKYKFODLDEESLKTQLAYFTDSKNTGR 915

RESULT 8

ABW01536

ID ABW01536 standard; protein; 1199 AA.

XX AC

XX ABW01536;

XX DT

15-JAN-2004 (first entry)

XX N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.
DE Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
KW N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.
XX Unidentified.
OS US2003124653-A1.
PN 03-JUL-2003.
XX 21-DEC-2001; 2001US-00023890.
XX 21-DEC-2001; 2001US-00023890.
PF (NOVA-) NOVAZYME PHARM INC.
XX Canfield WM;
PI WPI; 2003-810985/76.
DR N-PSDB; AAD62649.
XX Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX Claim 8; Page 10-13; 46pp; English.
XX The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing a glycoprotein
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
CC phosphotransferase protein
XX Sequence 1199 AA;

Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQVHVLDSYRDNIAGSKFQNRCLPMPIDVVTWNGTDLELLKELQOVRQMEBEQ 104
DB 35 SRDQVHVLDSYRDNIAGSKFQNRCLPMPIDVVTWNGTDLELLKELQOVRQMEBEQ 94

QY 105 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLDPALPANITLKDVPSPSPHSA 164
DB 95 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLDPALPANITLKDVPSPSPHSA 154

QY 165 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLM 224
DB 155 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLM 214

QY 225 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKQT 284
DB 215 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKQT 274

QY 285 KKNMTIDGKELTISPAYLLWDLISAISQSKQDEDIASRFEDNEELRYSLRSTIERHAPVR 344
DB 275 KKNMTIDGKELTISPAYLLWDLISAISQSKQDEDIASRFEDNEELRYSLRSTIERHAPVR 334

QY 345 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIY 404
DB 335 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIY 394

QY 405 LNDDVMFGKDWPDFFYSHSGKQKYLTPVNPNCAGCGPGSWIKDGYCDKACNNSACDWD 464
DB 395 LNDDVMFGKDWPDFFYSHSGKQKYLTPVNPNCAGCGPGSWIKDGYCDKACNNSACDWD 454

QY 465 GGDCSGNSGGSGRYIAGGGGTGSGIGVGPQWQGGGINSVSYCNQGCANSWLADKFCDOACN 524
DB 525 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGVEGAYSDNP 584
DB 515 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGVEGAYSDNP 574

QY 585 IIRHASIANKWKTHLIMHSGNATTHFNLTFTQNTDDEEFKMOITVEVDTRGPKLNST 644
DB 575 IIRHASIANKWKTHLIMHSGNATTHFNLTFTQNTDDEEFKMOITVEVDTRGPKLNST 634

QY 645 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRRRAQBEVKIPLVNIISLLP 704
DB 635 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRRRAQBEVKIPLVNIISLLP 694

QY 705 KDAQLSLNTLQLEHGDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSVA 764
DB 695 KDAQLSLNTLQLEHGDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSVA 754

QY 765 PQEKQVHKSILPNSILGVSERLQRLTFPAVSVKVNHGQGNPPLDLETTARFRVETHQK 824
DB 755 PQEKQVHKSILPNSILGVSERLQRLTFPAVSVKVNHGQGNPPLDLETTARFRVETHQK 814

QY 825 TIGGNVTKRPPSLIIVLESOMTEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHY 884
DB 815 TIGGNVTKRPPSLIIVLESOMTEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHY 874

QY 885 TDSYLGFLPWEKKYKFODLDEESLKTQLAYFTDSKNTGR 925
DB 875 TDSYLGFLPWEKKYKFODLDEESLKTQLAYFTDSKNTGR 915

```

Db 455 GDCSGNSGSRVYIAGGGTSGIGVGQPMQFGGINSVSYCNQGCANSWLADKFCDAQCN 514
Qy 525 VLSGCGDAGDCGDHPHLYKVIILLPNOHYIIPKGECLPYFSAFAVAKRGVEGAYSDNP 584
Db 515 VLSGCGDAGDCGDHPHLYKVIILLPNOHYIIPKGECLPYFSAFAVAKRGVEGAYSDNP 574
Qy 585 IIRHAIANKWTIHLIMHSGMNATIIHNLTFQNTNDEEFKMQITVEVDTRGPKLNST 644
Db 575 IIRHAIANKWTIHLIMHSGMNATIIHNLTFQNTNDEEFKMQITVEVDTRGPKLNST 634
Qy 645 AQGYENLVSPITLLPEABILFEDIPKEKRPKPKRHDVNSTRAQAEVKIPLVNIISLP 704
Db 635 AQGYENLVSPITLLPEABILFEDIPKEKRPKPKRHDVNSTRAQAEVKIPLVNIISLP 694
Qy 705 KDAQLSINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNQAIITDETNDLSVA 764
Db 695 KDAQLSINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNQAIITDETNDLSVA 754
Qy 765 PQEKQVHKILPNSLGVSRRLQRLTFPAVSVKNGHDOQNPPDLDTTARFRVETHQK 824
Db 755 PQEKQVHKILPNSLGVSRRLQRLTFPAVSVKNGHDOQNPPDLDTTARFRVETHQK 814
Qy 825 TIGGNTKPKPSLIVPLESQMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY 884
Db 815 TIGGNTKPKPSLIVPLESQMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY 874
Qy 885 TDSYLGFLPWEKKYFQDLDEESLKTQLAYFTDSKNTGR 925
Db 875 TDSYLGFLPWEKKYFQDLDEESLKTQLAYFTDSKNTGR 915

RESULT 9
AAE25294
ID AAE25294 standard; protein; 1196 AA.
XX
AC AAE25294;
DT
DT 30-OCT-2002 (first entry)
XX
DE Human nucleic acid-associated protein (NAAP-13).
XX
KW Human; nucleic acid-associated protein; NAAP-13; neurological disorder;
KW arteriosclerosis; cancer; cell proliferative disorder; arteriosclerosis;
KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
KW gene therapy; nootropic; neuroprotective; cerebroprotective; virucide;
KW immunosuppressive; protozoicide; antimicrobial.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide
FH /label= Signal_peptide
FT Domain
FT /note= "Cytosolic domain"
FT Domain
FT /note= "Transmembrane domain"
FT Protein
FT /note= "Mature human NAAP-13"
FT Domain
FT /note= "Non-cytosolic domain"
FT Domain
FT /note= "Transmembrane domain"
FT Domain
FT /note= "Cytosolic domain"
XX
PN WO200250279-A2.
XX
PD 27-JUN-2002.
XX
PF 19-DEC-2001; 2001WO-US050256.
XX
PP 21-DEC-2000; 2000US-0257714P.
XX

```

```

PR 05-JAN-2001; 2001US-0260081P.
PR 16-JAN-2001; 2001US-0262302P.
PR 23-JAN-2001; 2001US-0263823P.
PR 02-FEB-2001; 2001US-0266088P.
PR 29-OCT-2001; 2001US-0348442P.
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
XX Wallia NK, Tribouley KM, Yue H, Bacra S, Ding L, Lal PG;
PI Borowsky ML, Lu DM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
PI Gierzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
PI Lee EA, Yang J, Gorvad AB, Emerling BM, Marquis JP, Lee SY;
PI Swarnakar A, Reddy R;
XX WPI; 2002-519887/55.
DR N-PSDB; AAD41203.
XX
PT Nucleic acid associated proteins and nucleic acids for diagnosing,
PT treating and preventing cell proliferative (e.g. cancers), neurological
PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX
PS Claim 68; Page 169-172; 193pp; English.
XX
CC The invention relates to nucleic acid-associated proteins (NAAP) and
CC nucleic acids. The nucleic acid and amino acid sequences are useful for
CC diagnosing, treating and preventing cell proliferative e.g.
CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-13
XX
SQ Sequence 1196 AA;
Query Match 91.7%; Score 4500; DB 5; Length 1196;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 867; Conservative 1; Mismatches 0; Indels 60; Gaps 1;
Qy 1 MLFKLLQRTYTCLSHRYGLVYVGLVGVVTVISAFQGEVLEWSDQVHVFDSYRDNI 60
Db 1 MLFKLLQRTYTCLSHRYGLVYVGLVGVVTVISAFQGEVLEWSDQVHVFDSYRDNI 60
Qy 61 AGKSFQNRCLPMPIDIVYTWNGTDLLELKELOVREOMEEOEKAMRILGKNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDIVYTWNGTDLLELKELOVREOMEEOEKAMRILGKNTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPISLYPSFHSASDIFNVAKPKNSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPISLYPSFHSASDIFNVAKPKNSTNV 180
Qy 181 SVVVPDSTKQVEDAHSGLLKGNRQTVWRGYLTDDKEVPCLVLMQDLAFSLGPPPTFKET 240
Db 181 SVVVPDSTKQVEDAHSGLLKGNRQTVWRGYLTDDKEVPCLVLMQDLAFSLGPPPTFKET 240
Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNNMTIDGKELTISPA 300
Qy 301 YLLWDLISAISQSKODEDISASRFEDNEELRYLSURTERHAPWVRNIFVTNGQIPSWLNL 360
Db 301 YLLWDLISAISQSKODEDISASRFEDNEELRYLSURTERHAPWVRNIFVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFTYLLNDDVWFKDVPDDF 420
Db 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFTYLLNDDVWFKDVPDDF 420
Qy 421 YSHSKGQKQVLTWPVFNCAEGCGSWIKDGYCDKACNACDWDGDCSGNSGSRVYIAG 480
Db 421 YSHSKGQKQVLTWPVFNCAEGCGSWIKDGYCDKACNACDWDGDCSGNSGSRVYIAG 434

```

QY 481 CGGTSGIGVGHWPQFGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540
 Db 435 -----FGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 480
 QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600
 Db 481 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 540
 QY 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660
 Db 541 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 600
 QY 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
 Db 601 EAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 660
 QY 721 GDITLKGYNLSKALLRFLMNSQAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
 Db 661 GDITLKGYNLSKALLRFLMNSQAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 720
 QY 781 VSERLQRLTFPAVSVKVNGHDQGNPPDLLETTARFRVETHQTKTIGGNVTKKPPSLIV 840
 Db 721 VSERLQRLTFPAVSVKVNGHDQGNPPDLLETTARFRVETHQTKTIGGNVTKKPPSLIV 780
 QY 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900
 Db 781 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 840
 QY 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928
 Db 841 QDLDEESLKTQLAYFTDSKNTGRQLK 868

RESULT 10
 ADA54624
 ID AD54824 standard; protein; 846 AA.
 XX AC ADA54824;
 XX DT 20-NOV-2003 (first entry)
 XX DE Human protein, SEQ ID 2392.
 XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease.
 XX OS Homo sapiens.
 XX PN EP1293569-A2.
 XX PD 19-MAR-2003.
 XX PF 21-MAR-2002; 2002EP-00006586.
 XX PR 14-SEP-2001; 2001JP-00328381.
 XX FR 24-JAN-2002; 2002US-0350435P.
 XX PA (HELI-) HELIX RES INST.
 XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX WP1; 2003-395539/38.
 XX DR N-PSDB; ADA53185.
 XX PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX PS Claim 14; SEQ ID NO 2392; 205pp; English.

XX

The present invention relates to novel human secretory or membrane proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-ADA54071). The coding sequences are useful in the gene therapy of diseases caused by abnormalities of the proteins, e.g. cancer, inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 846 AA;

Query Match 91.1%; Score 4468; DB 6; Length 846;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIYSAQFGVLEWSDQVHVLFDSDYRDI 60

Db 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIYSAQFGVLEWSDQVHVLFDSDYRDI 60

QY 61 AGSFQNRJCLPMPIDVVTWNGTDLLELLKELQVREOMESEKAMREILGNVTEPTK 120

Db 61 AGSFQNRJCLPMPIDVVTWNGTDLLELLKELQVREOMESEKAMREILGNVTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLPALPANITLKDVPSPSYFHSASDIFNVAKPQNPSTNV 180

Db 121 KSEKQLECLLTHCIKVPMLVLPALPANITLKDVPSPSYFHSASDIFNVAKPQNPSTNV 180

QY 181 SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLMODLAFSLGPPPTFKET 240

Db 181 SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLMODLAFSLGPPPTFKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPNKDFQELNKQTKKNMTIDGKELTISPA 300

Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPNKDFQELNKQTKKNMTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRPEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360

Db 301 YLLWDLAISQSKQDEDISASRPEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFFSPAIESHYHRIEGLSQKFIYLLNDDVWFKQVWPDF 420

Db 361 DNPRVTIVTHQDVFRNLSHLPTFFSPAIESHYHRIEGLSQKFIYLLNDDVWFKQVWPDF 420

QY 421 YSHSGQKQVLTWPVPNCAGCGGKWDGYCDKACNNSACDWDGDCSGNSGGRYTAG 480

Db 421 YSHSGQKQVLTWPVPNCAGCGGKWDGYCDKACNNSACDWDGDCSGNSGGRYTAG 480

QY 481 GGTGSGIGVGHWPQFGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540

Db 481 GGTGSGIGVGHWPQFGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600

Db 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660

Db 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

Db 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDITLKGYNLSKALLRFLMNSQAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780

Db 721 GDITLKGYNLSKALLRFLMNSQAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780

QY 781 VSERLQRLTFPAVSVKVNGHDQGNPPDLLETTARFRVETHQTKTIGGNVTKKPPSLIV 840

Db 781 VSERLQRLTFPAVSVKVNGHDQGNPPDLLETTARFRVETHQTKTIGGNVTKKPPSLIV 840

QY 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900

Db 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900

RESULT 11
ADG39856
ID ADG39856 standard; protein; 847 AA.
XX AC ADG39856;
XX DT 26-FEB-2004 (first entry)
XX DE Protein similar to human NOV12 #2.
XX KW Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;
KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;
KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;
KW congenital adrenal hyperplasia; prostate cancer; diabetes;
KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;
KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;
KW multiple sclerosis; infectious disease; anorexia;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
XX KW wasting disorder.
XX OS Homo sapiens.
XX PN US2003203843-A1.
XX PD 30-OCT-2003.
XX PF 11-APR-2002; 2002US-00120801.
XX PR 20-APR-2001; 2001US-0285609P.
XX PR 23-APR-2001; 2001US-0285748P.
XX PR 24-APR-2001; 2001US-0286068P.
XX PR 25-APR-2001; 2001US-0286292P.
XX PR 03-MAY-2001; 2001US-0288334P.
XX PR 16-MAY-2001; 2001US-0291241P.
XX PR 14-SEP-2001; 2001US-0322284P.
XX PA (PENA)/ PENA C E A.
XX PA (GUOX)/ GUO X.
XX PA (SHIM)/ SHIMKETS R A.
XX PA (PADI)/ PADIGARU M.
XX PA (KEKU)/ KEKUDA R.
XX PA (SPYK)/ SPYTEK K A.
XX PA (MEHR)/ MEHRABAN F.
XX PA (TOPP)/ TOPPER J N.
XX PA (MALY)/ MALYANKAR U M.
XX PA (WASS)/ WASSERMAN S M.
XX PA (EDIN)/ EDINGER S R.
XX PA (SMIT)/ SMITHSON G.
XX PA (GUNT)/ GUNTHER E.
XX PA (KOMU)/ KOMUVES L.
XX PI PENA CEA, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA;
PI Mehraban F, Topper JN, Malvankar UM, Wasserman SM, Edinger SR;
PI Smithson G, Gunther E, Komuves L;
XX WP: 2003-900671/82.
XX New NOVX polypeptides and nucleic acids, useful for diagnosing or
PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,
PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or
PT multiple sclerosis.
XX PS Disclosure; SEQ ID NO 88; 215pp; English.
XX CC The invention relates to a new isolated polypeptide comprising an amino
CC acid sequence selected from 17 fully defined human NOVX sequences (even
CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX
CC amino acid or a variant of NOVX, where one or more amino acid residue in
CC the variant differs in no more than 15% of the amino acid residues of
CC NOVX. Also included are an isolated nucleic acid (NA) molecule
CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above

(odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment
encoding at least a portion of a NOVX polypeptide and a complement of
NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an
antibody that immunospecifically binds to NOVX, a method for determining
the presence or amount of NOVX or NOVX NA in a sample, a method of
identifying an agent that binds to NOVX, a method for identifying an
agent that modulates the expression or activity of NOVX, a method for
modulating the activity of NOVX, a method for screening for a modulator of
NOVX-associated disorder, a method for screening for a modulator of
activity or of latency or predisposition to a NOVX-associated disorder, a
method for determining the presence of or predisposition to a disease
associated with altered levels of NOVX or NOVX NA in a first mammalian
subject and a method of treating a pathological state in a mammal by
administering NOVX or an antibody that binds to NOVX. The NOVX
polypeptide, nucleic acid or antibody is useful in the manufacture of a
medicament for treating a syndrome associated with a human disease or a
NOVX-associated disorder. The NOVX polypeptides and nucleic acids
encoding them are useful for diagnosing or treating pathologies, diseases
or conditions associated with NOVX sequences, including cardiomyopathy,
atherosclerosis, hypertension, congenital heart defects, pulmonary
stenosis, scleroderma, obesity, metabolic disturbances associated with
obesity, transplantation, adrenoleukodystrophy, congenital adrenal
hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,
adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,
bronchial asthma, Crohn's disease, multiple sclerosis, infectious
disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,
or Parkinson's disease), immune disorders, haematopoietic disorders,
dyslipidaemias, and wasting disorders associated with chronic diseases
and as immunogens to produce antibodies to produce antibodies and as
vaccines. The sequences may further be used in chromosome mapping,
identifying individual from minute biological samples (tissue typing),
and in forensic identification of a biological sample. The present
sequence is a protein showing sequence similarity to a NOVX protein.

Sequence 847 AA;
Query Match 90.9%; Score 4461; DB 7; Length 847;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 842; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 MLFKLLQRTYTCLSHRYGLYVCFVGVVTVISAFQGEVLEWRSRQYHVLFDSDYRDI 60
Db 1 MLFKLLQRTYTCLSHRYGLYVCFVGVVTVISAFQGEVLEWRSRQYHVLFDSDYRDI 60
QY 61 AKGSFQNELCLPMPIDVYVTVWNGTDLLELKLQCVREQMEEEKAMREILGKNTTPTK 120
Db 61 AKGSFQNELCLPMPIDVYVTVWNGTDLLELKLQCVREQMEEEKAMREILGKNTTPTK 120
QY 121 KSEKQLECLLTHCIVKPMVLVLDPALPANITLKVPSLYPSPFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCIVKPMVLVLDPALPANITLKVPSLYPSPFHSASDIFNVAKPKNPSTNV 180
QY 181 SVVFPDSTKVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGCLVMDLAFILSGFPPTKET 240
Db 181 SVVFPDSTKVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGCLVMDLAFILSGFPPTKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKPDKQELNKPDKQELNKPDK 300
Db 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKPDKQELNKPDKQELNKPDK 300
QY 301 YLLWDLAISQSKODEDISASRFEDNEELRYSLRSIERHIEGLSQKFIVLNDVWPKDWDPDF 360
Db 301 YLLWDLAISQSKODEDISASRFEDNEELRYSLRSIERHIEGLSQKFIVLNDVWPKDWDPDF 360
QY 361 DNPRTVITHQDVFNRLSHLPTFSSPAIESHIEGLSQKFIVLNDVWPKDWDPDF 420
Db 361 DNPRTVITHQDVFNRLSHLPTFSSPAIESHIEGLSQKFIVLNDVWPKDWDPDF 420
QY 421 YSHSKGQKYLITWVPNCAEGCPGSIWKDGYCDKACNNACDWDGDCSGNSGSGRYIAG 480
Db 421 YSHSKGQKYLITWVPNCAEGCPGSIWKDGYCDKACNNACDWDGDCSGNSGSGRYIAG 480
QY 481 GGGTSGIGVHPWQFGGGINSVSYCNQCANSLADKFCQACNVLSGDFDAGDGGQDHF 540

Db 481 GGGTSGIGVGPWQGGGINSVYCNQGCANSLADKFCDCQACNVLSGFGDAGCGQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGNATTHFNLTFTONTNDEEPMQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660
Db 601 IMHSGNATTHFNLTFTONTNDEEPMQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPKEKPFKRDVNSTRAQAEVKIPLVNIISLLPKDAQLSNTLDLOLEH 720
Db 661 EAEILFEDIPKEKPFKRDVNSTRAQAEVKIPLVNIISLLPKDAQLSNTLDLOLEH 720
QY 721 GDITLKGYNLSKALLRFLMNSQHAQIKNOAITDETNDLSLVAPOKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRFLMNSQHAQIKNOAITDETNDLSLVAPOKQVHKSILPNSLG 780
QY 781 VSERLQRLTTPFAVSVKVNGHDQGNPPLDLETTARFRVETHQKTIGNVTKERPPSLIV 840
Db 781 VSERLQRLTTPFAVSVKVNGHDQGNPPLDLETTARFRVETHQKTIGNVTKERPPSLIV 840
QY 841 PLESQMT 847
Db 841 PLESQMT 847

RESULT 12

ABU07381 ID ABU07381 standard; protein; 1459 AA.

XX AC ABU07381;
XX XX

DT 28-JAN-2003 (first entry)

XX Human protein NOV12.

KW Human; NOX; cardiomyopathy; atherosclerosis; cancer; hypertension;
KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;
KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;
KW epilepsy; wasting disorder; neurogenesis; cell differentiation;
KW cell proliferation; haematopoiesis; wound healing; angiogenesis;
KW gene therapy; single nucleotide polymorphism; SNP.

XX Homo sapiens.

XX WO200285922-A2.

XX 31-OCT-2002.

XX 11-APR-2002; 2002WO-US011634.

XX 23-APR-2001; 2001US-0285748P.

XX 24-APR-2001; 2001US-0286068P.

XX 25-APR-2001; 2001US-0286292P.

XX 03-MAY-2001; 2001US-0288334P.

XX 16-MAY-2001; 2001US-0291241P.

XX 14-SEP-2001; 2001US-0322284P.

XX (CURA-) CURAGEN CORP.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pena CEA, Guo X, Shimkets RA, Padigar M, Kekuda R, Spytek KA;

XX Mehrahan F, Topper JN, Malyankar UM, Wasserman S, Edinger S;

XX Smithson G, Gunther E, Komuves L;

XX WPI; 2003-058712/05.

XX N-PSDB; ABX10235.

XX New NOX polypeptides and nucleic acids, useful for preventing or

PT treating NOX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
PS Claim 1; Page 94-95; 301pp; English.
XX
XX The invention relates to an isolated polypeptide comprising any one of 17
CC human NOX (1-9, 10a, 10b, 11-16) appearing as ABU07369-ABU07385, a
CC mature form of it, or a variant of them, where one or more residues of
CC the variant differs in not more than 15 % from the residues of the
CC sequence of them and their encoding polynucleotides appearing as ABX10223
CC ABX10239. Also included are NOX expression vectors, transformed cells,
CC antibodies, identifying an agent that binds to or modulates the
CC expression or activity of NOX and screening for a modulator of activity
CC or of latency or predisposition to a NOX-associated disorder. The NOX
CC polypeptides, polynucleotides and antibodies are useful in manufacturing
CC a medicament for treating or preventing a syndrome associated with NOX-
CC associated disorder, such as cardiomyopathy, atherosclerosis, cancer,
CC hypertension, diabetes, inflammation, autoimmune disorders, allergies,
CC blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,
CC asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis,
CC Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.
CC bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and
CC other wasting disorders associated with chronic diseases. The nucleic
CC acids and polypeptides may also be used as targets for the identification
CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC differentiation, cell proliferation, haematopoiesis, wound healing and
CC angiogenesis, in gene therapy, in generation of antibodies that bind
CC immunospecifically to NOX substances for use in therapeutic or
CC diagnostic methods. The nucleic acids are further used as hybridisation
CC probes, in chromosome mapping, tissue typing, preventive medicine, and
CC pharmacogenomics. The polypeptides are also useful as vaccines. The
CC present sequence represents a NOX protein
XX SQ Sequence 1459 AA;

Query Match 89.5%; Score 4391; DB 6; Length 1459;
Best Local Similarity 73.0%; Pred. No. 0;
Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;
QY 1 MLFKLLQRTYTCLSHRYGLYVCFILGVVTVISAFQFGE----- 39
Db 1 MLFKLLQRTYTCLSHRYGLYVCFILGVVTVISAFQFGEWEARDPAKHPVHRTAPT 60
QY 40 ----- 39
Db 61 NHPAQNVDSAEVKEGSGIRGKNGCRAVSLQDWPGRGCANFTFAFCHCKSEVSKRPL 120
QY 40 -----VYLEWSRDQYHVLFDSDYRD 58
Db 121 YILQNCWMLTDGWTWLLALLHGSLLIQGPASBPQCVLLKAKVLEWSRDQYHVLFDSDYRD 180
QY 59 NIAGKSFQNRCLCPMPIDVYVTVWNGTDLELLKELOQVREQMEBEQKAMRETLGNKNTTEP 118
Db 181 NIAGKSFQNRCLCPMPIDVYVTVWNGTDLELLKELOQVREQMEBEQKAMRETLGNKNTTEP 240
QY 119 TKKSEKQLECLLTHCICKVPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 178
Db 241 TKKSEKQLECLLTHCICKVPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 300
QY 179 NVSVVVFDSKTD-----VEDAHSLLKGN-----RQTVW----- 208
Db 301 NVSVVVFDSKTDGTLTLTKVTFEWKCEGEVASNANIGWTKDLSGSPRFLPMPVALEP 360
QY 209 -----RGYLTDDKEVPCVL-MODLAFSLGFP 234
Db 361 AQLSSALQILTFPRVSDRANTSVEIKLDTPLLRGYA---KPVGPEGLQPLFAHCLP 417
QY 235 -----PTF---KET-----NQLKT---KL 247
Db 418 TLDRKVNELRDFVKNYKQDPSILHKTCTFLREQVSMGESYVYKSEENIKELTKSKV 477
QY 248 PEN-----LSSKVKLLQLYS-----EASVALLKLNPKDFQELNKQTK 285

CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal
 CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,
 CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,
 CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious
 CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,
 CC or Parkinson's disease), immune disorders, haematopoietic disorders,
 CC dyslipidaemias, and wasting disorders associated with chronic diseases.
 CC The polypeptides can be used as immunogens to produce antibodies and as
 CC vaccines. The sequences may further be used in chromosome mapping,
 CC identifying individual from minute biological samples (tissue typing),
 CC and in forensic identification of a biological sample. The present
 CC sequence represents a NOVX protein.

XX SQ Sequence 1459 AA;

Query Match 89.5%; Score 4391; DB 7; Length 1459;

Best Local Similarity 73.0%; Pred. No. 0;

Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;

QY 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVTVTSVFQGE----- 39
 DB 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVTVTSVFQGEWVEARDPAKHPVHRTATTK 60
 QY 40 ----- 39
 DB 61 NHPAQNVDSEAEKSGIRRGKNGCRVSLQDPWPGTRCANFTAFCHDKCFSEVSQKRF 120
 QY 40 -----VLEWRSRDQVHVLFDSDYRD 58
 DB 121 YILQCHWLTDGWTWLLALHGSLLLOGPASEPCVLLKAKVLEWSDRDQVHVLFDSDYRD 180
 QY 59 NIAGKSFONRLCLPMPIDVVTVWNGTDLLELLKLOVQREOMEEOEKAMEIILGKNTTEP 118
 DB 181 NIAGKSFONRLCLPMPIDVVTVWNGTDLLELLKLOVQREOMEEOEKAMEIILGKNTTEP 240
 QY 119 TKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 178
 DB 241 TKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 300
 QY 179 NVSVVVFDSKTD-----VEDAHSGLLKGN-----RQTW----- 208
 DB 301 NVSVVVFDSKTDGTLTKQVTFEWEKCEGEVASNANIGKTDLGSPERPLPWPVLEPPR 360
 QY 209 -----RGYTTDKVEPGLV-NQDLAFISGFP 234
 DB 361 AOLSSALQILTRPVSDRANTSVEIKLDTPLLAGYA---KVPFGPETGLQPLSFARCLP 417
 QY 235 -----PTF---XET-----NOLKT---KL 247
 DB 418 TDLRKYNELRDFVMYKQDPSILHTKTCFLREQVSMGESYKSEENIKELTKGKV 477
 QY 248 PEN-----LSKVKLLQLYS-----EASVALLKNNPKDFQELNKKQTK 285
 DB 478 EENISTDELASEEDLEIDNEAVIEPTDTSPOEMGDGEASVALLKNNPKDFQELNKKQTK 537
 QY 286 KNMTIDGKELTISPAVLLWDLTSAISQSKQDEDISASFEDNEELRYSLRSIERHAPWVRN 345
 DB 538 KNMTIDGKELTISPAVLLWDLTSAISQSKQDEDISASFEDNEELRYSLRSIERHAPWVRN 597
 QY 346 IFIVTNGQIFSWLNDLPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSKQFIYL 405
 DB 598 IFIVTNGQIFSWLNDLPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSKQFIYL 657
 QY 406 NDDVMFGKDVDPDFYSHSGKQKYLTPVPNCAEGCPGSKIWDGCDKACNNACSDWDG 465
 DB 658 NDDVMFGKDVDPDFYSHSGKQKYLTPVPNCAEGCPGSKIWDGCDKACNNACSDWDG 717
 QY 466 GDCSGNSGGSGRYTAGGGGTGSGIVGHPWQFGGGINSVSYCNQGCANSLWADKFCDAQNV 525
 DB 718 GDCSGNSGGSGRYTAGGGGTGSGIVGHPWQFGGGINSVSYCNQGCANSLWADKFCDAQNV 777
 QY 526 LSCGFDAGDCGQ-----DHPHELYKVILLPNQTHYII 557

DB 778 LSCGFDAGDCGQENSDSKNRKTEKCPVKKKIMFLFFPLDHPHELYKVILLPNQTHYII 837
 QY 558 PKGECPLPYTFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHNLT 617
 DB 838 PKGECPLPYTFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHNLT 857
 QY 618 QNTNDEEFKQIITVEVDTRREGPKLNSTAKGVENLVSPITLLPEABILFEDIPKEKRF 677
 DB 898 QNTNDEEFKQIITVEVDTRREGPKLNSTAKGVENLVSPITLLPEABILFEDIPKEKRF 957
 QY 678 FKEDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNTLDLQLEHGDITLKGYNLSKALL 737
 DB 958 FKEDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNTLDLQLEHGDITLKGYNLSKALL 1017
 QY 738 SFLMNSCHAKIKNOAIITDETNDLSLVAPOEKVHKSIPLNSLGVSRLOQLTFPVSVKV 797
 DB 1018 SFLMNSCHAKIKNOAIITDETNDLSLVAPOEKVHKSIPLNSLGVSRLOQLTFPVSVKV 1077
 QY 798 NGHDQONPDLDTTARFRVETHOKTIGGNVTKEKPPSLVPLESOMTKKKTGK 857
 DB 1078 NGHDQONPDLDTTARFRVETHOKTIGGNVTKEKPPSLVPLESOMTKKKTGK 1137
 QY 858 ENSRMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYFQDLDDEESLKTQAYF 917
 DB 1138 ENSRMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYFQDLDDEESLKTQAYF 1197
 QY 918 TDSKNTGROLK 928
 DB 1198 TDSKNTGROLK 1208

RESULT 14
 ABR61384
 ID ABR61384 standard; protein; 908 AA.
 XX AC ABR61384;
 XX DT 01-AUG-2003 (first entry)
 XX DE Murine GlcNAc-phosphotransferase alpha-subunit.
 XX KW Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
 KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
 KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
 KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
 KW lysosomal storage disease; enzyme; alpha-subunit.
 XX OS Mus musculus.
 XX PN US6537785-B1.
 XX PD 25-MAR-2003.
 XX PF 10-AUG-2000; 2000US-00636077.
 XX PR 14-SEP-1999; 99US-0153831P.
 XX PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.
 XX PI Canfield WW;
 XX DR WPI; 2001-290356/30.
 XX DR N-PSDB; ACC81007.
 XX PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.
 XX PS Disclosure; Page 44-47; 62pp; English.
 XX CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester

CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (ii). The protein of the invention has
 CC nephrotropic activity, and may be useful in enzyme replacement therapy. A
 CC protein of the invention (i), (ii) is useful for preparing a
 CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
 CC comprising a terminal mannose-6-phosphate, is useful for treating a
 CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention
 XX
 SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 7.2e-305;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQFGEVLEWSRDQYHVLFDSDYRNI 60
 DB 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQFGEVLEWSRDQYHVLFDSDYRNI 60
 QY 61 AGKSFQNLCLPMPIDVVYTWNGTDLLELKELOQVREOQEEQKAMREILGKNTTEPTK 120
 DB 61 AGKSFQNLCLPMPIDVVYTWNGTDLLELKELOQVREOQEEQKAMREILGKNTTEPTK 120
 QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYSFHSGASDIFNVAKPKNPSTNV 180
 DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYSFHSGASDIFNVAKPKNPSTNV 180
 QY 181 SVVVFSTKDVDEAHSGLLKGNSTWRGVLTTDKAPGLVLIQGLAFSGFPPTFKET 240
 DB 181 SVVVFSTKDVDEAHSGLLKGNSTWRGVLTTDKAPGLVLIQGLAFSGFPPTFKET 240
 QY 241 NQTKLKPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNTIDGKELTISP 299
 DB 241 NQTKLKPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNTIDGKELTISP 299

QY 300 AYLLMDLSAISQSKQDEEDSASHFEDNEELRYLSIERHAPKVRNIFIVTNGQIIPSWLN 359
 DB 301 AYLLMDLSAISQSKQDEEDSASHFEDNEELRYLSIERHAPKVRNIFIVTNGQIIPSWLN 360
 QY 360 LDNPRVITVHQDVFRLNLSLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPD 419
 DB 361 LDNPRVITVHQDVFRLNLSLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPD 420
 QY 420 FYSHSKGQKVLTPWPCNACGPGSWIKGYCDKACNSACDWDGDCSGNSGGRYIA 479
 DB 421 FYSHSKGQKVLTPWPCNACGPGSWIKGYCDKACNTSPCDWDGDCSGNTAGNRFA 480
 QY 480 GGGTGSTGVGHWPQFGGINSVSYCNQGCANSWLADKFCQACNVLSGCFDAGDCGQDH 539
 DB 481 RGGTGNIGAGQHWQFGGINTSYCNQGCANSWLADKFCQACNVLSGCFDAGDCGQDH 540
 QY 540 PHELYKVILLPNQTHYIIPKGECLUPYFSAEVAKRGVEGAYSNDPIIRHASIANKWKTH 599
 DB 541 PHELYKVILLPNQTHYIIPKGECLUPYFSAEVAKRGVEGAYSNDPIIRHASIANKWKTH 600
 QY 600 LIMSGMNTTIIHNLIFQNTNDEEFKQITVEVDTEGPKLNSTAOQYENLVSPITLL 659
 DB 601 LIMSGMNTTIIHNLIFQNTNDEEFKQITVEVDTEGPKLNSTAOQYENLVSPITLL 660
 QY 660 PEAEILFEDIPKEKFPKSHDVNSTRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLE 719
 DB 661 PQADVPEDEVKKEKFPKSHDVNSTRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLE 720
 QY 720 HGDITLKYNLSKALLSFLMNSQAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSL 779
 DB 721 RGDITLKYNLSKALLSFLMNSQAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSL 777
 QY 780 GVSRERLQRLTFPVSVKVNGHQQONPFLDLETTARFRVETHQTGKTGGNVTKKPPSLI 839
 DB 778 AGEHRSERWTAPAEVTIVKGRHALNPPVPLETNARL----AQPTLGVTVSKENISPLI 832
 QY 840 VPLSQMTKEKKTGKEKENERMEENAHNGVTEVLLGRKLQHYTDSYLGFLPWEKKY 899

DB 833 VPESHLP-----KEESDRAEGNA---VPVKELVPGRRLO---QNPVGFPLWEKKY 879
 QY 900 FQDLDEEESIKTOLAYFTDSKNTGROLK 928
 DB 980 FQDLDEEESIKTOLAYFTDRKHTGROLK 908

RESULT 15
 ADD27817
 ID ADD27817 standard; protein; 908 AA.
 XX
 AC ADD27817;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE GlcNAc-phosphotransferase associated protein #1.
 XX
 KW mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
 KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.
 XX
 OS Mus musculus.
 XX
 PN US2003119088-A1.
 XX
 PD 26-JUN-2003.
 XX
 PF 21-DEC-2001; 2001US-00023888.
 XX
 PR 21-DEC-2001; 2001US-00023888.
 XX
 PA (NOVA-) NOVAYME PHARM INC.
 XX
 PI Canfield W, Kudo M;
 XX
 DR WPI; 2003-801323/75.
 XX
 PT Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.
 XX
 PS Disclosure; SEQ ID NO 9; 55pp; English.
 XX
 CC The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of a GlcNAc-phosphotransferase associated protein.
 XX
 SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;
 Best Local Similarity 79.9%; Pred. No. 7.2e-305;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQFGEVLEWSRDQYHVLFDSDYRNI 60
 DB 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQFGEVLEWSRDQYHVLFDSDYRNI 60
 QY 61 AGKSFQNLCLPMPIDVVYTWNGTDLLELKELOQVREOQEEQKAMREILGKNTTEPTK 120
 DB 61 AGKSFQNLCLPMPIDVVYTWNGTDLLELKELOQVREOQEEQKAMREILGKNTTEPTK 120
 QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYSFHSGASDIFNVAKPKNPSTNV 180
 DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYSFHSGASDIFNVAKPKNPSTNV 180
 QY 181 SVVVFSTKDVDEAHSGLLKGNSTWRGVLTTDKAPGLVLIQGLAFSGFPPTFKET 240
 DB 181 PVVVFSTKDVDEAHSGLLKGNSTWRGVLTTDKAPGLVLIQGLAFSGFPPTFKET 240
 QY 241 NQTKLKPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNTIDGKELTISP 299

[illegible]

Search completed: November 21, 2004, 12:54:44
Job time : 112.597 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:47:18 ; Search time 27.9723 Seconds
(without alignments)
2200.144 Million cell updates/sec

Title: US-10-023-888-4
Perfect score: 4907
Sequence: 1 MLFKLQRTYTCLSHRYGL.....SLXTQLAYFTDKNTRQLX 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5E.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4	US-09-635-872A-1
2	4907	100.0	928	4	US-09-636-077A-1
3	4907	100.0	928	4	US-09-636-060C-1
4	4907	100.0	928	4	US-09-986-552-1
5	4907	100.0	928	4	US-09-636-596C-1
6	3894	79.4	908	4	US-09-635-872A-15
7	3894	79.4	908	4	US-09-636-077A-15
8	3894	79.4	908	4	US-09-636-060C-15
9	3894	79.4	908	4	US-09-986-552-15
10	3894	79.4	908	4	US-09-636-596C-15
11	503	10.3	113	4	US-09-635-872A-11
12	503	10.3	113	4	US-09-636-077A-11
13	503	10.3	113	4	US-09-636-060C-11
14	503	10.3	113	4	US-09-986-552-11
15	503	10.3	113	4	US-09-636-596C-11
16	488	9.9	391	4	US-09-270-767-42309
17	479	9.8	502	4	US-09-635-872A-13
18	479	9.8	502	4	US-09-636-077A-13
19	479	9.8	502	4	US-09-636-060C-13
20	479	9.8	502	4	US-09-986-552-13
21	479	9.8	502	4	US-09-636-596C-13
22	265	5.4	545	4	US-08-936-107A-10
23	255.5	5.2	546	4	US-08-311-731A-248
24	240	4.9	364	4	US-09-328-352-5382
25	168.5	3.4	1015	1	US-08-537-210A-1
26	168.5	3.4	1015	3	US-09-113-825-1
27	168.5	3.4	2471	1	US-08-185-432-16

28	168.5	3.4	2471	1	US-08-083-590A-19	Sequence 19, Appl
29	168.5	3.4	2471	3	US-08-532-384-19	Sequence 19, Appl
30	168.5	3.4	2471	4	US-08-899-232-1	Sequence 1, Appl
31	168.5	3.4	2471	4	US-09-121-457-1	Sequence 1, Appl
32	154.5	3.1	2321	4	US-09-230-652-2	Sequence 2, Appl
33	152	3.1	1964	3	US-09-467-997-1	Sequence 1, Appl
34	145.5	3.0	1068	1	US-08-537-210A-2	Sequence 2, Appl
35	145.5	3.0	1068	3	US-09-113-825-2	Sequence 2, Appl
36	145.5	3.0	2556	1	US-08-185-432-17	Sequence 17, Appl
37	145.5	3.0	2556	3	US-08-083-590A-20	Sequence 20, Appl
38	145.5	3.0	2556	3	US-08-532-384-20	Sequence 20, Appl
39	145.5	3.0	2556	4	US-08-899-232-2	Sequence 2, Appl
40	145.5	3.0	2556	4	US-09-121-457-2	Sequence 2, Appl
41	141	2.9	1078	1	US-08-264-534-32	Sequence 32, Appl
42	141	2.9	1078	1	US-08-083-590A-11	Sequence 11, Appl
43	141	2.9	1078	1	US-08-465-500-32	Sequence 32, Appl
44	141	2.9	1078	2	US-08-346-128-32	Sequence 32, Appl
45	141	2.9	1078	3	US-08-532-384-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-635-872A-1
; Sequence 1, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE.
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-635-872A-1

Query Match		100.0%	Score 4907;	DB 4;	Length 928;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 928;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLFKLQRTYTCLSHRYGLYVCFLGVVTVIVSAFQGEVWLEWSDQVHVLFDSDYRDN	60		
DB	1	MLFKLQRTYTCLSHRYGLYVCFLGVVTVIVSAFQGEVWLEWSDQVHVLFDSDYRDN	60		
QY	61	AGKSFQNRICLPMPIDVYVTVWNGTDLLELLKLOQVREOMEEOQKMRILGKNTTEPTK	120		
DB	61	AGKSFQNRICLPMPIDVYVTVWNGTDLLELLKLOQVREOMEEOQKMRILGKNTTEPTK	120		
QY	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPSTNV	180		
DB	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPSTNV	180		
QY	181	SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTQVKEVGLVMDLAFLSGFPPTFKET	240		
DB	181	SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTQVKEVGLVMDLAFLSGFPPTFKET	240		
QY	241	NQLKTLPNLSKVKLLQLYSEASVALLKNNPKDFQELNKQTKNMTIDGKELTISPA	300		
DB	241	NQLKTLPNLSKVKLLQLYSEASVALLKNNPKDFQELNKQTKNMTIDGKELTISPA	300		
QY	301	YLLWDLAISQSKQDEDISASFEONEELRYSLRSTIERHAPVWRNIFIVTNGQIPSWLNL	360		
DB	301	YLLWDLAISQSKQDEDISASFEONEELRYSLRSTIERHAPVWRNIFIVTNGQIPSWLNL	360		
QY	361	DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHRIEGLSKFYLNDVDFGKDVDPDF	420		

Db 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQFYIYLNDDVMFGKDVWPPDF 420
 QY 421 YSHSGQKVYLTWPVNCACGPGSWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480
 Db 421 YSHSGQKVYLTWPVNCACGPGSWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480
 QY 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540
 Db 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540
 QY 541 HELYKVILLPNOHYIIPKGECLPYFSAEVAKGVGAYSNDNPIIRHASIAKWKTIHL 600
 Db 541 HELYKVILLPNOHYIIPKGECLPYFSAEVAKGVGAYSNDNPIIRHASIAKWKTIHL 600
 QY 601 IMHSGMNATTHFNLTQNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
 Db 601 IMHSGMNATTHFNLTQNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
 QY 661 EAEILFEDIPEKFPKRDVNSTRAQBEVKIPLVNIISLIPKDAQLSLNTLDLQLEH 720
 Db 661 EAEILFEDIPEKFPKRDVNSTRAQBEVKIPLVNIISLIPKDAQLSLNTLDLQLEH 720
 QY 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
 Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
 QY 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840
 Db 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840
 QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 Db 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 QY 901 QDLLDEESLKTQLAYFTDSKNTGRQLK 928
 Db 901 QDLLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 2
 US-09-636-077A-1
 ; Sequence 1, Application US/09636077A
 ; Patent No. 6537785
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS OF TREATING LYSSOMAL STORAGE DISEASE
 ; FILE REFERENCE: 19612US0
 ; CURRENT APPLICATION NUMBER: US/09/636,077A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 928
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-636-077A-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRQYTCLSHRYGLYVCFGLGVVVTVISAFQGEVVLWSRDQYHVLPSYRDI 60
 Db 1 MLFKLLQRQYTCLSHRYGLYVCFGLGVVVTVISAFQGEVVLWSRDQYHVLPSYRDI 60
 QY 61 AGKSFQRLCLPMPIDVVYTWANGTDLLELKELOQVREQMEBEQKAMREILGNTEPTK 120
 Db 61 AGKSFQRLCLPMPIDVVYTWANGTDLLELKELOQVREQMEBEQKAMREILGNTEPTK 120
 QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180

Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180
 QY 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFISGFPPTFKET 240
 Db 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFISGFPPTFKET 240
 QY 241 NOLKTKLPENISSKVKLLQLLYSEASVALLKLNPKDPQELNKTQKMTIDGKELTISPA 300
 Db 241 NOLKTKLPENISSKVKLLQLLYSEASVALLKLNPKDPQELNKTQKMTIDGKELTISPA 300
 QY 301 YLLWDLISAISQKODEDIASRPFEDNEELRYSLSIERHAPWVRNIFIVINGQIPSWLNL 360
 Db 301 YLLWDLISAISQKODEDIASRPFEDNEELRYSLSIERHAPWVRNIFIVINGQIPSWLNL 360
 QY 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQFYIYLNDDVMFGKDVWPPDF 420
 Db 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQFYIYLNDDVMFGKDVWPPDF 420
 QY 421 YSHSGQKVYLTWPVNCACGPGSWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480
 Db 421 YSHSGQKVYLTWPVNCACGPGSWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480
 QY 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540
 Db 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540
 QY 541 HELYKVILLPNOHYIIPKGECLPYFSAEVAKGVGAYSNDNPIIRHASIAKWKTIHL 600
 Db 541 HELYKVILLPNOHYIIPKGECLPYFSAEVAKGVGAYSNDNPIIRHASIAKWKTIHL 600
 QY 601 IMHSGMNATTHFNLTQNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
 Db 601 IMHSGMNATTHFNLTQNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
 QY 661 EAEILFEDIPEKFPKRDVNSTRAQBEVKIPLVNIISLIPKDAQLSLNTLDLQLEH 720
 Db 661 EAEILFEDIPEKFPKRDVNSTRAQBEVKIPLVNIISLIPKDAQLSLNTLDLQLEH 720
 QY 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
 Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
 QY 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840
 Db 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840
 QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 Db 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
 QY 901 QDLLDEESLKTQLAYFTDSKNTGRQLK 928
 Db 901 QDLLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 3
 US-09-636-060C-1
 ; Sequence 1, Application US/09636060C
 ; Patent No. 6642038
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM M
 ; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSSOMAL TARGETING PATHWAY
 ; FILE REFERENCE: 210119USOCONT
 ; CURRENT APPLICATION NUMBER: US/09/636,060C
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 928
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-09-636-060C-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRYGLVYVCFGLGVVTVIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60
DB 1 MLFKLQRTYTCLSHRYGLVYVCFGLGVVTVIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREQMEBEQKAMREILGKNTTPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREQMEBEQKAMREILGKNTTPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPSTNV 180

QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240

QY 241 NQKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300
DB 241 NQKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360
DB 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHIERHAPWVRNIFIVTNGQIPSWNL 420
DB 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHIERHAPWVRNIFIVTNGQIPSWNL 420

QY 421 YSHSGQKQVLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYIAG 480
DB 421 YSHSGQKQVLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYIAG 480

QY 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCQACNVLSGCGFDAGCGQDHF 540
DB 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCQACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660
DB 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDTILKGNLSKALLRSFLMNSOHAQIKNOAIIITDETNDLSLVAPOEQVHKSILPNSLG 780
DB 721 GDTILKGNLSKALLRSFLMNSOHAQIKNOAIIITDETNDLSLVAPOEQVHKSILPNSLG 780

QY 781 VSELRQRLTPAVSKVNGHDQGNPPLDLETTARFVETHQKTIGNVTKEKPPSLIV 840
DB 781 VSELRQRLTPAVSKVNGHDQGNPPLDLETTARFVETHQKTIGNVTKEKPPSLIV 840

QY 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
DB 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900

QY 901 QDLADEESLKTQLAYFTDSKNTGRQLK 928
DB 901 QDLADEESLKTQLAYFTDSKNTGRQLK 928

RESULT 4

US-09-986-552-1

; Sequence 1, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE;
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-986-552-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRYGLVYVCFGLGVVTVIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60
DB 1 MLFKLQRTYTCLSHRYGLVYVCFGLGVVTVIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREQMEBEQKAMREILGKNTTPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELLEKQVREQMEBEQKAMREILGKNTTPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPSTNV 180

QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240

QY 241 NQKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300
DB 241 NQKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360
DB 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHIERHAPWVRNIFIVTNGQIPSWNL 420
DB 361 DNPRTVIVTHQDVFRNLSHLPTFSSPAIESHIERHAPWVRNIFIVTNGQIPSWNL 420

QY 421 YSHSGQKQVLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYIAG 480
DB 421 YSHSGQKQVLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYIAG 480

QY 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCQACNVLSGCGFDAGCGQDHF 540
DB 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCQACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660
DB 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDTILKGNLSKALLRSFLMNSOHAQIKNOAIIITDETNDLSLVAPOEQVHKSILPNSLG 780

721 GDITLKGYNLSKALLRSFLNLSQAHAKIKNOAIIITDETNSLVAPOEKQVHKSILPNSLG 780
781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKKPPSLIV 840
781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKKPPSLIV 840
841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928
901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928
RESULT 5
US-09-636-596C-1
; Sequence 1, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLCNase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-596C-1
Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLFKLLQRTQTYTCLSHRYGLVVCFLGVVTVTSFAQFGEVLEWSDQYHVLFDSDYRNI 60
DB 1 MLFKLLQRTQTYTCLSHRYGLVVCFLGVVTVTSFAQFGEVLEWSDQYHVLFDSDYRNI 60
QY 61 AGKSFQNRKLCPLMPIDVVYTWVNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRKLCPLMPIDVVYTWVNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVDLPALPANITLKVPSLYPSFHSASDIPNVAKPNSNV 180
DB 121 KSEKQLECLLTHCIKVPMLVDLPALPANITLKVPSLYPSFHSASDIPNVAKPNSNV 180
QY 181 SVVVFDSKDYEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMDLAFSLGPPPTFKET 240
DB 181 SVVVFDSKDYEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMDLAFSLGPPPTFKET 240
QY 241 NQKTKLPENISSKVKLLQLYSEASVALLKNNPKQFQELNKMOTDKGKELTISP 300
DB 241 NQKTKLPENISSKVKLLQLYSEASVALLKNNPKQFQELNKMOTDKGKELTISP 300
QY 301 YLLWDLISAISOKQDEDSASREFNEDNELRSLRSIERHAPVWNIIVINGQIPSWNL 360
DB 301 YLLWDLISAISOKQDEDSASREFNEDNELRSLRSIERHAPVWNIIVINGQIPSWNL 360
QY 361 DNPRTVIVTHQDVFNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVNFQKDWPDFF 420
DB 361 DNPRTVIVTHQDVFNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVNFQKDWPDFF 420
QY 421 YSHSGKQKYLTPVPNCACGCPGSIWKDGYCDKACNNSACDWDGCGSGNSGSRVAG 480
DB 421 YSHSGKQKYLTPVPNCACGCPGSIWKDGYCDKACNNSACDWDGCGSGNSGSRVAG 480
QY 481 GGGTSGIGVGHFWQFGGGINSVSYCNOGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540

481 GGGTSGIGVGHFWQFGGGINSVSYCNOGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540
541 HELYKIVLLPNQTHYIIPKGECLPYFSAFAVAKGVEGAYSDNPRIIRHASIANKWKTIHL 600
541 HELYKIVLLPNQTHYIIPKGECLPYFSAFAVAKGVEGAYSDNPRIIRHASIANKWKTIHL 600
601 IMHSGMAATTHFNLTONTNDEBFKQITVEVDTRGPKLNSTAOAGYENLVSPIITLLP 660
601 IMHSGMAATTHFNLTONTNDEBFKQITVEVDTRGPKLNSTAOAGYENLVSPIITLLP 660
661 EAEILFEDIPKEKPPKFRHDVNSTRAQBEVKIPLVNISSLPLPKDAQLSINTLDLQEH 720
661 EAEILFEDIPKEKPPKFRHDVNSTRAQBEVKIPLVNISSLPLPKDAQLSINTLDLQEH 720
721 GDITLKGYNLSKALLRSFLMNSOAHAKIKNOAIIITDETNSLVAPOEKQVHKSILPNSLG 780
721 GDITLKGYNLSKALLRSFLMNSOAHAKIKNOAIIITDETNSLVAPOEKQVHKSILPNSLG 780
781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKKPPSLIV 840
781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKKPPSLIV 840
841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928
901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928
RESULT 6
US-09-635-872A-15
; Sequence 15, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 1956113US
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-15
Query Match 79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;
QY 1 MLFKLLQRTQTYTCLSHRYGLVVCFLGVVTVTSFAQFGEVLEWSDQYHVLFDSDYRNI 60
DB 1 MLFKLLQRTQTYTCLSHRYGLVVCFLGVVTVTSFAQFGEVLEWSDQYHVLFDSDYRNI 60
QY 61 AGKSFQNRKLCPLMPIDVVYTWVNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRKLCPLMPIDVVYTWVNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVDLPALPANITLKVPSLYPSFHSASDIPNVAKPNSNV 180
DB 121 KSEKQLECLLTHCIKVPMLVDLPALPANITLKVPSLYPSFHSASDIPNVAKPNSNV 180
QY 181 SVVVFDSKDYEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMDLAFSLGPPPTFKET 240
DB 181 SVVVFDSKDYEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMDLAFSLGPPPTFKET 240
QY 241 NQKTKLPENISSKVKLLQLYSEASVALLKNNPKQFQELNKMOTDKGKELTISP 299

```
Db 241 SOLKTKLPKAPFKIKILRLYSEASVALLKNNPKGQELNKKTKKNTIDGKELTISP 300
QY 300 AYLLMDLSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVRNFIPIVINGQIPSWLN 359
Db 301 AYLLMDLSAISOSKODEDASASRFEDNEELRYSLRSIERHAPWVRNFIPIVINGQIPSWLN 360
QY 360 LNDPRVTIVTHQDVPRNLSHLPTSSPAIESHRIEGLSOKFYLYLNDNDVMFGKDVWDD 419
Db 361 LNDPRVTIVTHQDVPRNLSHLPTSSPAIESHRIEGLSOKFYLYLNDNDVMFGKDVWDD 420
QY 420 FYSHSGQKQVLTWPVPCABGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGSRYYA 479
Db 421 FYSHSGQKQVLTWPVPCABGCGPSWIKDGYCDKACNNSPCDMDGDCSGNTAGNFVA 480
QY 480 GGGTGSIGVGHWPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCDGDCGDH 539
Db 481 RGGGTGNIAGQHWQFGGINTISYCNOCGANSWLADKFCDOACNVLSGCDGDCGDH 540
QY 540 FHELYKVILLPNQTHYVVPKGEYLSYFSFANIARKRIEGTYSNDNPIIRHASIANKWKTLH 599
Db 541 FHELYKVILLPNQTHYVVPKGEYLSYFSFANIARKRIEGTYSNDNPIIRHASIANKWKTLH 600
QY 600 LIMHSGMNTATTHFNLTQNTNDEEFKMOITVEVDTRGPKLNSTAKQGYENLVSPITLL 659
Db 601 LIMPGMNATTIYFNLTQNDANDEEFKIQIAVEVDTRGPKLNSTAKQGYENLVSPITPL 660
QY 660 PEABILFEDIPEKRPFKRHDVNSTRAOEKVIPLVNIISLLPKDAQLSLNLDLQLE 719
Db 661 PQADVPEFEDVPEKRPFKIRHDVNTGRFQEVKIPRVNIIISLLPKDAQVLSNLDLQLE 720
QY 720 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSLIPNSL 779
Db 721 RGDITLKGYNLSKALLRSFLGNSLDTKIKPQA-RTDETKGNLEVPQENPSHR--PHGF 777
QY 780 GVSERLQRLTTPPAVSVKYNHGDQGNPPDLDETTARFVRVETHQKTIGNVTKKPSLI 839
Db 778 AGEHRSERWTAPAEITVVKGRDHALNPPPVLETNARL-----AQPTLGVTVSKENLSPLI 832
QY 840 VPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLQHYTDSYGLFLEWKKY 899
Db 833 VPPESHLP-----KEESDRAEGNA---VPKELVPGRLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
Db 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908
```

RESULT 7

US-09-636-077A-15
; Sequence 15, Application US/09636077A

; Patent No. 6537785

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS OF TREATING LYSSOMAL STORAGE DISEASE

; FILE REFERENCE: 195612US0

; CURRENT APPLICATION NUMBER: US/09/636,077A

; CURRENT FILING DATE: 2000-08-10

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 15

; LENGTH: 908

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-636-077A-15

Query Match

Best Local Similarity 79.4%; Score 3894; DB 4; Length 908;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVVTIVSAFQFGEVVLWSRQDQYHVLFDSDYRDN 60

```
Db 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVVTIVSAFQFGEVVLWSRQDQYHVLFDSDYRDN 60
QY 61 AGKSFQNRCLPMPIDVVTWNGTDLLELLKELQQVREOMESEBQKAMREILGNKNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDVVTWNGTDLLELLKELQQVREOMESEBQKAMREILGNKNTTEPTK 120
QY 121 KSEKQLECLLTHCHICKYKPMVLVDPALPANITLKDVPSPLYSPHSASDIPNVAKPKNSTNV 180
Db 121 KSEKQLECLLTHCHICKYKPMVLVDPALPANITLKDVPSPLYSPHSASDIPNVAKPKNSTNV 180
QY 181 SVVVFSTKDVEDAHAGLLKGNRSQRTVWRGYLTDDKEVPLGLVLMQDLAFLSGPPPTFKET 240
Db 181 PVVVFSTKDVEDAHAGPFGKGOQTDVWRAYLTTDKDAPGLVLIQGLAFLSGPPPTFKET 240
QY 241 NQKTKLPEN-LSSVKYLLQLYSEASVALLKNNPKDFOELNKKTKKNTIDGKELTISP 299
Db 241 SQKTKLPKAPFKIKILRLYSEASVALLKNNPKGQELNKKTKKNTIDGKELTISP 300
QY 300 AYLLMDLSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVRNFIPIVINGQIPSWLN 359
Db 301 AYLLMDLSAISOSKODEDASASRFEDNEELRYSLRSIERHAPWVRNFIPIVINGQIPSWLN 360
QY 360 LNDPRVTIVTHQDVPRNLSHLPTSSPAIESHRIEGLSOKFYLYLNDNDVMFGKDVWDD 419
Db 361 LNDPRVTIVTHQDVPRNLSHLPTSSPAIESHRIEGLSOKFYLYLNDNDVMFGKDVWDD 420
QY 420 FYSHSGQKQVLTWPVPCABGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGSRYYA 479
Db 421 FYSHSGQKQVLTWPVPCABGCGPSWIKDGYCDKACNNSPCDMDGDCSGNTAGNFVA 480
QY 480 GGGTGSIGVGHWPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCDGDCGDH 539
Db 481 RGGGTGNIAGQHWQFGGINTISYCNOCGANSWLADKFCDOACNVLSGCDGDCGDH 540
QY 540 FHELYKVILLPNQTHYVVPKGEYLSYFSFANIARKRIEGTYSNDNPIIRHASIANKWKTLH 599
Db 541 FHELYKVILLPNQTHYVVPKGEYLSYFSFANIARKRIEGTYSNDNPIIRHASIANKWKTLH 600
QY 600 LIMHSGMNTATTHFNLTQNTNDEEFKMOITVEVDTRGPKLNSTAKQGYENLVSPITLL 659
Db 601 LIMPGMNATTIYFNLTQNDANDEEFKIQIAVEVDTRGPKLNSTAKQGYENLVSPITPL 660
QY 660 PEABILFEDIPEKRPFKRHDVNSTRAOEKVIPLVNIISLLPKDAQLSLNLDLQLE 719
Db 661 PQADVPEFEDVPEKRPFKIRHDVNTGRFQEVKIPRVNIIISLLPKDAQVLSNLDLQLE 720
QY 720 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSLIPNSL 779
Db 721 RGDITLKGYNLSKALLRSFLGNSLDTKIKPQA-RTDETKGNLEVPQENPSHR--PHGF 777
QY 780 GVSERLQRLTTPPAVSVKYNHGDQGNPPDLDETTARFVRVETHQKTIGNVTKKPSLI 839
Db 778 AGEHRSERWTAPAEITVVKGRDHALNPPPVLETNARL-----AQPTLGVTVSKENLSPLI 832
QY 840 VPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLQHYTDSYGLFLEWKKY 899
Db 833 VPPESHLP-----KEESDRAEGNA---VPKELVPGRLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
Db 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908
```

RESULT 8

US-09-636-060C-15

; Sequence 15, Application US/09636060C

; Patent No. 6642038

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSSOMAL TARGETING PATHWAY

; FILE REFERENCE: 21019US0CONT

; CURRENT APPLICATION NUMBER: US/09/636,060C

```
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-15

Query Match      79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

Qy 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQFGEVLEWSRDQYHVLFDSDYRDN1 60
Db 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQFGEVLEWSRDQYHVLFDSDYRDN1 60
Qy 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLEKQVREMEEQKAMREILGKNTTEPTK 120
Db 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLEKQVREMEEQKAMREILGKNTTEPTK 120
Qy 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLEKQVREMEEQKAMREILGKNTTEPTK 120
Db 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLEKQVREMEEQKAMREILGKNTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPYPSFHSASDIFNVAKPNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPYPSFHSASDIFNVAKPNPSTNV 180
Qy 181 SVVVPDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240
Db 181 SVVVPDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240
Qy 181 PVVVPDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240
Db 181 PVVVPDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240
Qy 241 NOLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKQNTIDGKELTISP 299
Db 241 NOLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKQNTIDGKELTISP 299
Qy 241 SOLTKLPKAPPLKILRLYSEASVALLKNNPKDFQELNKOTKQNTIDGKELTISP 300
Db 241 SOLTKLPKAPPLKILRLYSEASVALLKNNPKDFQELNKOTKQNTIDGKELTISP 300
Qy 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVVRNIPIVINGQIPSWLN 359
Db 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVVRNIPIVINGQIPSWLN 359
Qy 301 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVVRNIPIVINGQIPSWLN 360
Db 301 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVVRNIPIVINGQIPSWLN 360
Qy 360 LONPRVITVTHQDVPFNLISLPTSSPAIESHRIEGLSOKFIYLNDDVDFKDVWDD 419
Db 360 LONPRVITVTHQDVPFNLISLPTSSPAIESHRIEGLSOKFIYLNDDVDFKDVWDD 419
Qy 361 LONPRVITVTHQDVPFNLISLPTSSPAIESHRIEGLSOKFIYLNDDVDFKDVWDD 420
Db 361 LONPRVITVTHQDVPFNLISLPTSSPAIESHRIEGLSOKFIYLNDDVDFKDVWDD 420
Qy 420 FVSHSGKQKYLITWVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
Db 420 FVSHSGKQKYLITWVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
Qy 421 FVSHSGKQKYLITWVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 480
Db 421 FVSHSGKQKYLITWVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 480
Qy 480 GGGTGSIGVGHWPQFQGGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
Db 480 GGGTGSIGVGHWPQFQGGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
Qy 481 RGGTGNIGAGCHQWQFQGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
Db 481 RGGTGNIGAGCHQWQFQGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
Qy 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVGEYSNDPIIRHASIANKWKTH 599
Db 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVGEYSNDPIIRHASIANKWKTH 599
Qy 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVGEYSNDPIIRHASIANKWKTH 600
Db 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVGEYSNDPIIRHASIANKWKTH 600
Qy 600 LIMSGMNATTHHNLFQNTNDEEFKQITVEVDREGKLNASTAGKYNLVSPTLL 659
Db 600 LIMSGMNATTHHNLFQNTNDEEFKQITVEVDREGKLNASTAGKYNLVSPTLL 659
Qy 601 LIMPGMNATTHYFNLTQNDSEFKIQIAVEVDTEAPKLNSTTQKAYESLVSPVTL 660
Db 601 LIMPGMNATTHYFNLTQNDSEFKIQIAVEVDTEAPKLNSTTQKAYESLVSPVTL 660
Qy 660 PEASILPEDIPEKRFPPKRDVNSRRQAEVKEIPLVNIISLIPKDAQSLNLTDLQLE 719
Db 660 PEASILPEDIPEKRFPPKRDVNSRRQAEVKEIPLVNIISLIPKDAQSLNLTDLQLE 719
Qy 661 PQADVPEDEVPEKRFPPKRDVNSRRQAEVKEIPLVNIISLIPKDAQSLNLTDLQLE 720
Db 661 PQADVPEDEVPEKRFPPKRDVNSRRQAEVKEIPLVNIISLIPKDAQSLNLTDLQLE 720
Qy 720 HGDITLKGYNLSKALLERFLMNSQHAKIRNQAIITDENDSLVAPQEKVHKSILPNSL 779
Db 720 HGDITLKGYNLSKALLERFLMNSQHAKIRNQAIITDENDSLVAPQEKVHKSILPNSL 779
Qy 721 RGDITLKGYNLSKALLERFLMNSQHAKIRNQAIITDENDSLVAPQEKVHKSILPNSL 777
Db 721 RGDITLKGYNLSKALLERFLMNSQHAKIRNQAIITDENDSLVAPQEKVHKSILPNSL 777
Qy 780 GVSRELQRLTPAVSVKNGHQDQONPLDLETTARFVETHQTKTGGNVTEKEPPLI 839
Db 780 GVSRELQRLTPAVSVKNGHQDQONPLDLETTARFVETHQTKTGGNVTEKEPPLI 839
Qy 778 AGEHSEKWTAPATVTVKGRDHALNPPVLETNARL-----AQPTLGVTVSKENLSPLI 832
Db 778 AGEHSEKWTAPATVTVKGRDHALNPPVLETNARL-----AQPTLGVTVSKENLSPLI 832
Qy 840 VPLESQMTKEKKTGCKEKNRMBENAEHNGVTEVLLGRKLQHYTDSYLGFPWEKKY 899
Db 840 VPLESQMTKEKKTGCKEKNRMBENAEHNGVTEVLLGRKLQHYTDSYLGFPWEKKY 899
```

QY 600 LIMHSGMNAITTHNLTFQNTNDEBFKMOITVEVDTRGPKLNSTAKGYENLVSPITLL 659
DB 601 LIMPGGMNATTIYNLTQNDANDEBFKQIAVEVDTRAPKLNSTTQKAYESLVSPVTP 660
QY 660 PEAEILFEDIPEKRPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQLSNTLDLQLE 719
DB 661 PQADVPEFDPKRPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQLSNTLDLQLE 720
QY 720 HGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 779
DB 721 RGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 777
QY 780 GVSERLQRLTTPAVSVKYNHGDQGNPPDLLETTARFRVETHTQKTGGNVTKKPPSLI 839
DB 778 AGEHRSERWTAPAEVTTVKGRDHALNPPVLETNARL-----AQPTLGTVTSKENLSPLI 832
QY 840 VPLESOMTKKKTGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKY 899
DB 833 VPESHLP-----KEESDRAEGNA---VPVKELVPGRRLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
DB 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908

RESULT 10

US-09-636-596C-15

; Sequence 15, Application US/09636596C

; Patent No. 6770468

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY

; FILE REFERENCE: 10929-0001-77

; CURRENT APPLICATION NUMBER: US/09/636.596C

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 15

; LENGTH: 908

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-636-596C-15

Query Match

Best Local Similarity 79.9%; Score 3894; DB 4; Length 908;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLQROTYTCISHRVGLVVCFLGVVTVTSVAFQGEVVLKSRDQYHVLFDSDYDNI 60
DB 1 MLFKLQROTYTCISHRVGLVVCFLGVVTVTSVAFQGEVVLKSRDQYHVLFDSDYDNI 60
QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLEKQVREEMEEQKAMREILGKNTTPTK 120
DB 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLEKQVREEMEEQKAMREILGKNTTPTK 120
QY 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNSTNV 180
DB 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNSTNV 180
QY 181 SVVVFDSKQVEDAHSGLLKGNSTQTVMRGYLTDDKVPGLVLMODLAFSLGPPPTPKET 240
DB 181 PVVVFDTKQVEDAHAGFPKGGQGTQVWRAVLTDDKAPGLVLIQGLAFSLGPPPTPKET 240
QY 241 NQTKTKLPEN-LSKVKLLQLYSEASVALLKNNPKDFQELNKQTKQNTIDGKELTISP 299
DB 241 SOLTKTKLPKAPFLKIKLLRLYSEASVALLKNNPKGFQELNKQTKQNTIDGKELTISP 300
QY 300 AYLLWDLISATOSKODEDISASRFEDNEELRYSLRSTERHAPVVRNFIIVTNGQIPSWLN 359
DB 301 AYLLWDLISATOSKODEDISASRFEDNEELRYSLRSTERHAPVVRNFIIVTNGQIPSWLN 360

QY 360 LDNPRVTIYTHODVFNLSHLPTFFSPATESHRIEGLSOKFIYLNDDVMFGKWDPD 419
DB 361 LDNPRVTIYTHODVFNLSHLPTFFSPATESHRIEGLSOKFIYLNDDVMFGKWDPD 420
QY 420 FYSHSGQKQVYLTWVPNCAEGCPGSKIWDGCDKACNNSACDWDGDCGSGSGGSRVIA 479
DB 421 FYSHSGQKQVYLTWVPNCAEGCPGSKIWDGCDKACNNSPCDWDGCGSGSGTAGNRFVA 480
QY 480 GGGGTGASIGVHPWQGGGGINSVSYCNCANSLWADKFCDOACNVLSGFGDAGCGQDH 539
DB 481 RGGGTGNIAGGHWQGGGINTISYCNOCANSLWADKFCDOACNVLSGFGDAGCGQDH 540
QY 540 FHELYKVILLPQOYTHIIPKGECLPYFSAEVAKEVGEAGVSDNPPIRHASIANKWKTH 599
DB 541 FHELYKVILLPQOYTHIIPKGECLPYFSAEVAKEVGEAGVSDNPPIRHASIANKWKTH 600
QY 600 LIMHSGMNAITTHNLTFQNTNDEBFKMOITVEVDTRGPKLNSTAKGYENLVSPITLL 659
DB 601 LIMPGGMNATTIYNLTQNDANDEBFKQIAVEVDTRAPKLNSTTQKAYESLVSPVTP 660
QY 660 PEAEILFEDIPEKRPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQLSNTLDLQLE 719
DB 661 PQADVPEFDPKRPFKRHDVNSTRAQEEVKIPLVNI:SLLPKDAQLSNTLDLQLE 720
QY 720 HGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 779
DB 721 RGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 777
QY 780 GVSERLQRLTTPAVSVKYNHGDQGNPPDLLETTARFRVETHTQKTGGNVTKKPPSLI 839
DB 778 AGEHRSERWTAPAEVTTVKGRDHALNPPVLETNARL-----AQPTLGTVTSKENLSPLI 832
QY 840 VPLESOMTKKKTGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKY 899
DB 833 VPESHLP-----KEESDRAEGNA---VPVKELVPGRRLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
DB 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908

RESULT 11

US-09-635-872A-11

; Sequence 11, Application US/09635872A

; Patent No. 6534300

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 195613US0

; CURRENT APPLICATION NUMBER: US/09/635,872A

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-635-872A-11

Query Match

Best Local Similarity 10.3%; Score 503; DB 4; Length 113;

Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 FPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKQTKQNTIDG 292
DB 1 FPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKGFQELNKQTKQNTIDG 60
QY 293 KELTISPALLWDLISATOSKODEDISASRFEDNEELRYSLRSTERH 339
DB 61 KELTISPALLWDLISATOSKODEDISASRFEDNEELRYSLRSTERH 107

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-986-552-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.2e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDG 292
Db 1 FPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNMSISG 60
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

RESULT 15
US-09-636-596C-11
; Sequence 11, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-CLONACASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-636-596C-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.2e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDG 292
Db 1 FPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNMSISG 60
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

Search completed: November 21, 2004, 13:03:15
Job time : 30.9723 secs

RESULT 12
US-09-636-077A-11
; Sequence 11, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612USO
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-636-077A-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.2e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDG 292
Db 1 FPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNMSISG 60
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

RESULT 13
US-09-636-060C-11
; Sequence 11, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 21019USOCNT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-636-060C-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.2e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDG 292
Db 1 FPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNMSISG 60
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

RESULT 14
US-09-986-552-11
; Sequence 11, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:56:09 ; Search time 100.171 Seconds
(without alignments)
3280.691 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907

Sequence: 1 MLFKLLQRTYTCLSHRYGL.....SLKQLAYFTDKNTGRLK 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score Greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4907	100.0	928	9	US-09-895-072-1
2	4907	100.0	928	9	US-09-986-552-1
3	4907	100.0	928	14	US-10-023-888-4
4	4907	100.0	928	14	US-10-023-889-4
5	4907	100.0	928	14	US-10-023-890-4
6	4907	100.0	928	14	US-10-024-197-4
7	4907	100.0	928	14	US-10-023-894-4
8	4907	100.0	928	14	US-10-306-686-1
9	4630	94.4	1199	14	US-10-023-888-2
10	4630	94.4	1199	14	US-10-023-889-2
11	4630	94.4	1199	14	US-10-023-890-2
12	4630	94.4	1199	14	US-10-024-197-2
13	4630	94.4	1199	14	US-10-023-894-2

```

14 4473 91.2 847 14 US-10-120-801-88 Sequence 88, Appl
15 4468 91.1 846 14 US-10-094-749-2392 Sequence 2392, Ap
16 4391 89.5 1459 14 US-10-120-801-26 Sequence 26, Appl
17 3894 79.4 908 9 US-09-895-072-15 Sequence 15, Appl
18 3894 79.4 908 9 US-09-986-552-15 Sequence 15, Appl
19 3894 79.4 908 14 US-10-023-888-9 Sequence 9, Appl
20 3894 79.4 908 14 US-10-023-889-9 Sequence 9, Appl
21 3894 79.4 908 14 US-10-023-890-9 Sequence 9, Appl
22 3894 79.4 908 14 US-10-024-197-9 Sequence 9, Appl
23 3894 79.4 908 14 US-10-023-894-9 Sequence 9, Appl
24 3894 79.4 908 14 US-10-306-686-15 Sequence 15, Appl
25 2021 41.2 663 9 US-10-120-801-87 Sequence 87, Appl
26 1900 38.7 367 9 US-09-864-761-42893 Sequence 42893, A
27 1900 38.7 367 14 US-10-023-386-32927 Sequence 32927, A
28 1292 26.3 384 14 US-10-139-794-173 Sequence 173, Appl
29 734.5 15.0 384 14 US-10-120-801-89 Sequence 89, Appl
30 503 10.3 113 9 US-09-895-072-11 Sequence 11, Appl
31 503 10.3 113 9 US-09-986-552-11 Sequence 11, Appl
32 503 10.3 113 14 US-10-023-888-14 Sequence 14, Appl
33 503 10.3 113 14 US-10-023-889-14 Sequence 14, Appl
34 503 10.3 113 14 US-10-023-890-14 Sequence 14, Appl
35 503 10.3 113 14 US-10-024-197-14 Sequence 14, Appl
36 503 10.3 113 14 US-10-023-894-14 Sequence 14, Appl
37 503 10.3 113 14 US-10-306-686-11 Sequence 11, Appl
38 488 9.9 652 14 US-10-120-801-91 Sequence 91, Appl
39 479 9.8 502 9 US-09-895-072-13 Sequence 13, Appl
40 479 9.8 502 9 US-09-986-552-13 Sequence 13, Appl
41 479 9.8 502 14 US-10-023-888-16 Sequence 16, Appl
42 479 9.8 502 14 US-10-023-889-16 Sequence 16, Appl
43 479 9.8 502 14 US-10-023-890-16 Sequence 16, Appl
44 479 9.8 502 14 US-10-024-197-16 Sequence 16, Appl
45 479 9.8 502 14 US-10-023-894-16 Sequence 16, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-895-072-1
; Sequence 1. Application US/09895072
; Patent No. US20020025550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-895-072-1

```

Query Match 100.0%; Score 4907; DB 9; Length 928;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVVCFLGVVVVTVSAFQFGEVLEWSDQYHVLFDSDYRNI 60

Db 1 MLFKLLQRTYTCLSHRYGLVVCFLGVVVVTVSAFQFGEVLEWSDQYHVLFDSDYRNI 60

QY 61 AGKSFQNLCLPMPIDVVTYTWNGTDLLELKLQVREQMEEQKAMREILGKNTTEPTK 120

Db 61 AGKSFQNLCLPMPIDVVTYTWNGTDLLELKLQVREQMEEQKAMREILGKNTTEPTK 120

QY 121 KSEQLECLLTHCIVKPMVLDPALPANTLKDVPSLPFSFASDIFNVAKPNPSTNV 180

|||||

Db 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPSTNV 180
QY 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKEVPGVLVMDLAFSLGFPPTFKET 240
Db 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKEVPGVLVMDLAFSLGFPPTFKET 240
QY 241 NQKTKLPENLSKVLLQLYSEASVALLKNNPKDFQELNKOTKNNWIDGKELTISPA 300
Db 241 NQKTKLPENLSKVLLQLYSEASVALLKNNPKDFQELNKOTKNNWIDGKELTISPA 300
QY 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360
Db 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360
QY 361 DNPRVTIVTHQDVFNRLSHLPTSSPAIESHIIHRIEGLSQKFIYLNDDVWFGKDVWDDF 420
Db 361 DNPRVTIVTHQDVFNRLSHLPTSSPAIESHIIHRIEGLSQKFIYLNDDVWFGKDVWDDF 420
QY 421 YSHSKGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
Db 421 YSHSKGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIVGHPWQFGGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Db 481 GGGTSGIVGHPWQFGGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
QY 541 HELYKVILLNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTREGPKLNSTAGKYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTREGPKLNSTAGKYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRPFKRHVDVNSTRAOEVEKIPLVNISLPKDAQSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKRPFKRHVDVNSTRAOEVEKIPLVNISLPKDAQSLNTLDLQLEH 720
QY 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
QY 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840
Db 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840
QY 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDDEESLKTQLAYFTDTSKNTGRQLK 928
Db 901 QDLDDEESLKTQLAYFTDTSKNTGRQLK 928

RESULT 2
US-09-986-552-1
; Sequence 1, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US/7DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 928

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-1
Query Match 100.0%; Score 4907; DB 9; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLFLLQRYTCLSHRYGLYVCFGLVWVTIVSAFQFGEVWLEWSDQHVLFDSYRDN 60
Db 1 MLFLLQRYTCLSHRYGLYVCFGLVWVTIVSAFQFGEVWLEWSDQHVLFDSYRDN 60
QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLELLKELQVREQMEBECKAMREILGNTEPTK 120
Db 61 AGKSFQNRCLPMPIDVVYTWNGTDLELLKELQVREQMEBECKAMREILGNTEPTK 120
QY 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPSTNV 180
QY 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKEVPGVLVMDLAFSLGFPPTFKET 240
Db 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKEVPGVLVMDLAFSLGFPPTFKET 240
QY 241 NQKTKLPENLSKVLLQLYSEASVALLKNNPKDFQELNKOTKNNWIDGKELTISPA 300
Db 241 NQKTKLPENLSKVLLQLYSEASVALLKNNPKDFQELNKOTKNNWIDGKELTISPA 300
QY 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360
Db 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWNL 360
QY 361 DNPRVTIVTHQDVFNRLSHLPTSSPAIESHIIHRIEGLSQKFIYLNDDVWFGKDVWDDF 420
Db 361 DNPRVTIVTHQDVFNRLSHLPTSSPAIESHIIHRIEGLSQKFIYLNDDVWFGKDVWDDF 420
QY 421 YSHSKGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
Db 421 YSHSKGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIVGHPWQFGGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Db 481 GGGTSGIVGHPWQFGGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
QY 541 HELYKVILLNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTREGPKLNSTAGKYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTQNTNDEEFKQITVEVDTREGPKLNSTAGKYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRPFKRHVDVNSTRAOEVEKIPLVNISLPKDAQSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKRPFKRHVDVNSTRAOEVEKIPLVNISLPKDAQSLNTLDLQLEH 720
QY 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
QY 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840
Db 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840
QY 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDDEESLKTQLAYFTDTSKNTGRQLK 928
Db 901 QDLDDEESLKTQLAYFTDTSKNTGRQLK 928

RESULT 3

US-10-023-888-4
; Sequence 4, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-4

Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDI 60
DB 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDI 60

QY 61 AKGSFQNRCLCLPMPIDVVTWVNGTDLLELLEKQVREOMEEBQKAMREILGKNTTEPTK 120
DB 61 AKGSFQNRCLCLPMPIDVVTWVNGTDLLELLEKQVREOMEEBQKAMREILGKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180

QY 181 SVVVPDSTKDVDAHSGLLKGNRSQTVMRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240
DB 181 SVVVPDSTKDVDAHSGLLKGNRSQTVMRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240

QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360
DB 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDVWPDF 420
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDVWPDF 420

QY 421 YSHSGKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480
DB 421 YSHSGKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480

QY 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540
DB 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVLIPNQTHTYIIPKGCLEPFGFAEYAKGVGAYSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKVLIPNQTHTYIIPKGCLEPFGFAEYAKGVGAYSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPTILP 660
DB 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPTILP 660

QY 661 EAEILFEDIPEKEFPFKRHVDVNSTRAQBEVKIPLVNTSLLPKDAQLSINTLDLQLEH 720
DB 661 EAEILFEDIPEKEFPFKRHVDVNSTRAQBEVKIPLVNTSLLPKDAQLSINTLDLQLEH 720

QY 721 GDITLKGYNLSKALLSFLMNSQAKIKNOAITDETNDLSVAPQEKVHKSTLPSNLG 780
DB 721 GDITLKGYNLSKALLSFLMNSQAKIKNOAITDETNDLSVAPQEKVHKSTLPSNLG 780

RESULT 4

US-10-023-889-4
; Sequence 4, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARB
; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-889-4

Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDI 60
DB 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDI 60

QY 61 AKGSFQNRCLCLPMPIDVVTWVNGTDLLELLEKQVREOMEEBQKAMREILGKNTTEPTK 120
DB 61 AKGSFQNRCLCLPMPIDVVTWVNGTDLLELLEKQVREOMEEBQKAMREILGKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180

QY 181 SVVVPDSTKDVDAHSGLLKGNRSQTVMRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240
DB 181 SVVVPDSTKDVDAHSGLLKGNRSQTVMRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240

QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360
DB 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDVWPDF 420
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFIYLNDDVMFGKDVWPDF 420

QY 421 YSHSGKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480
DB 421 YSHSGKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480

QY 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540
DB 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGMNAITTHFNLTQNTNDEBFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
DB 601 IMHSGMNAITTHFNLTQNTNDEBFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
QY 721 GDITLKGYNLSKALLRSFLMNSOHAQKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
DB 721 GDITLKGYNLSKALLRSFLMNSOHAQKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
QY 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFVETHTQKTIGGNVTKEKPPSLIV 840
DB 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFVETHTQKTIGGNVTKEKPPSLIV 840
QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
DB 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLLEBESLKTQLAYFTDSKNTGRQLK 928
DB 901 QDLLEBESLKTQLAYFTDSKNTGRQLK 928

RESULT 5

US-10-023-890-4
; Sequence 4, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-4

Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60
DB 1 MLFKLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60
QY 61 AKGSFQNLCLPMPIDVVYTWNGTDLLEKELQVREOQMEEBEOKAMREILGKNTTEPTK 120
DB 61 AKGSFQNLCLPMPIDVVYTWNGTDLLEKELQVREOQMEEBEOKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHC1KVPMLVLDPALPANITLKVPSIYPSFHSASDIFNVAKPNPSTNV 180
DB 121 KSEKQLECLLTHC1KVPMLVLDPALPANITLKVPSIYPSFHSASDIFNVAKPNPSTNV 180
QY 181 SVVVFSDTKVDSDAHSGLKNSRQTVWRGYLTDDKEVPGVLVMDLAFSLSGPPPTFKET 240
DB 181 SVVVFSDTKVDSDAHSGLKNSRQTVWRGYLTDDKEVPGVLVMDLAFSLSGPPPTFKET 240
QY 241 NQLXTKLPELSSKVKLLQLYSEASVALLKNNPKDFQELNKTQKNTIDGKELTISPA 300
DB 241 NQLXTKLPELSSKVKLLQLYSEASVALLKNNPKDFQELNKTQKNTIDGKELTISPA 300
QY 301 YLLWDLISAISQKODEISASRPFEDNEELRYLSRSIERHAPVWRNIPIVINGQIPSWLNL 360

DB 301 YLLWDLISAISQKODEISASRPFEDNEELRYLSRSIERHAPVWRNIPIVINGQIPSWLNL 360
QY 361 DNPRTVITVTHQDVFNLSHLPTFSSPAIESHIHRIEGLSQFIYLNDDVNFVFGDWWPDDF 420
DB 361 DNPRTVITVTHQDVFNLSHLPTFSSPAIESHIHRIEGLSQFIYLNDDVNFVFGDWWPDDF 420
QY 421 YSHSGQKQVYLITWVPNCAEGCPGSIWKDGYCDKACNNSACDMDGDCSGNSGGSRYIAG 480
DB 421 YSHSGQKQVYLITWVPNCAEGCPGSIWKDGYCDKACNNSACDMDGDCSGNSGGSRYIAG 480
QY 481 GGGTSGICGVHPWPGGGINSVSYCNOGCANSWLADKFCQACNVLSGCFDAGCGQDHF 540
DB 481 GGGTSGICGVHPWPGGGINSVSYCNOGCANSWLADKFCQACNVLSGCFDAGCGQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGMNAITTHFNLTQNTNDEBFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
DB 601 IMHSGMNAITTHFNLTQNTNDEBFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPEKEFPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
QY 721 GDITLKGYNLSKALLRSFLMNSOHAQKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
DB 721 GDITLKGYNLSKALLRSFLMNSOHAQKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
QY 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFVETHTQKTIGGNVTKEKPPSLIV 840
DB 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFVETHTQKTIGGNVTKEKPPSLIV 840
QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
DB 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLLEBESLKTQLAYFTDSKNTGRQLK 928
DB 901 QDLLEBESLKTQLAYFTDSKNTGRQLK 928

RESULT 6

US-10-024-197-4
; Sequence 4, Application US/10024197
; Publication No. US2003013924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS C
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-4

Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60
DB 1 MLFKLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60
QY 61 AKGSFQNLCLPMPIDVVYTWNGTDLLEKELQVREOQMEEBEOKAMREILGKNTTEPTK 120

```
Db 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREIILGKNTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Qy 181 SVVVFSTKDVDEAHSGLLKGNRSQTVMRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Db 181 SVVVFSTKDVDEAHSGLLKGNRSQTVMRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Qy 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLHLPTFFSPPAIESHIERIEGLSQFIYLNDDVMFGKDVWPDF 420
Db 361 DNPRVTIVTHQDVFRNLHLPTFFSPPAIESHIERIEGLSQFIYLNDDVMFGKDVWPDF 420
Qy 421 YSHSGQKQVYLTWPVPCNAEGCPGSIKDGKCDKACNNSACDWDGDCSGNSGGSRYIAG 480
Db 421 YSHSGQKQVYLTWPVPCNAEGCPGSIKDGKCDKACNNSACDWDGDCSGNSGGSRYIAG 480
Qy 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGGDHF 540
Db 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNAITTHFNLTQNTNDEEFKQIITVEVDTREGPKLNSTAQKGYENLVSPITLLP 660
Db 601 IMHSGMNAITTHFNLTQNTNDEEFKQIITVEVDTREGPKLNSTAQKGYENLVSPITLLP 660
Qy 661 EAEILFEDIPKEKPPFKRHVDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Db 661 EAEILFEDIPKEKPPFKRHVDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Qy 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Qy 781 VSERLQRLTTPAVSVKVNQHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSLIV 840
Db 781 VSERLQRLTTPAVSVKVNQHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900

RESULT 7
US-10-023-894-4
; Sequence 4, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; TITLE OF INVENTION: ACETYLGLUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDAS
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-4

Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFKLQRTYTCLSHRYGLYVCFGLGVVVTIISAFQFGEVLEWSDQHVHVFDSYRDI 60
Db 1 MLFKLQRTYTCLSHRYGLYVCFGLGVVVTIISAFQFGEVLEWSDQHVHVFDSYRDI 60
Qy 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREIILGKNTTEPTK 120
Db 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREIILGKNTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPYPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPYPSFHSASDIFNVAKPKNPSTNV 180
Qy 181 SVVVFSTKDVDEAHSGLLKGNRSQTVMRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Db 181 SVVVFSTKDVDEAHSGLLKGNRSQTVMRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Qy 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLHLPTFFSPPAIESHIERIEGLSQFIYLNDDVMFGKDVWPDF 420
Db 361 DNPRVTIVTHQDVFRNLHLPTFFSPPAIESHIERIEGLSQFIYLNDDVMFGKDVWPDF 420
Qy 421 YSHSGQKQVYLTWPVPCNAEGCPGSIKDGKCDKACNNSACDWDGDCSGNSGGSRYIAG 480
Db 421 YSHSGQKQVYLTWPVPCNAEGCPGSIKDGKCDKACNNSACDWDGDCSGNSGGSRYIAG 480
Qy 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGGDHF 540
Db 481 GGGTSGIGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNAITTHFNLTQNTNDEEFKQIITVEVDTREGPKLNSTAQKGYENLVSPITLLP 660
Db 601 IMHSGMNAITTHFNLTQNTNDEEFKQIITVEVDTREGPKLNSTAQKGYENLVSPITLLP 660
Qy 661 EAEILFEDIPKEKPPFKRHVDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Db 661 EAEILFEDIPKEKPPFKRHVDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Qy 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Qy 781 VSERLQRLTTPAVSVKVNQHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSLIV 840
Db 781 VSERLQRLTTPAVSVKVNQHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Qy 901 QDLDEESLKTQLAYFTDSKNTGROLK 928
Db 901 QDLDEESLKTQLAYFTDSKNTGROLK 928
```

RESULT 8
 US-10-306-686-1
 ; Sequence 1, Application US/10306686
 ; Publication No. US20030148460A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: PHOSPHOTRANSFER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY
 ; FILE REFERENCE: 230397US77D1V
 ; CURRENT APPLICATION NUMBER: US/10/306,686
 ; CURRENT FILING DATE: 2002-11-29
 ; PRIOR APPLICATION NUMBER: 09/636,596
 ; PRIOR FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-08-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 928
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-306-686-1

Query Match 100.0%; Score 4907; DB 14; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLFKLLQRTYTCLSHRYGLYVCFGLVVTIVSAFQFGEVVLWMSRDQYHVLFDSDYRNI	60
DB	1	MLFKLLQRTYTCLSHRYGLYVCFGLVVTIVSAFQFGEVVLWMSRDQYHVLFDSDYRNI	60
QY	61	AGKSFQRLCLPMPIDVYTWNGTDLLEKKELOQVREOEEQKAMREILGNKTEPTK	120
DB	61	AGKSFQRLCLPMPIDVYTWNGTDLLEKKELOQVREOEEQKAMREILGNKTEPTK	120
QY	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSYFHSASDIFNVAKPNPSTNV	180
DB	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSYFHSASDIFNVAKPNPSTNV	180
QY	181	SVVVFSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVMDQLAFSLGPPPTFKET	240
DB	181	SVVVFSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVMDQLAFSLGPPPTFKET	240
QY	241	NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQTKNNMTIDGKELTISPA	300
DB	241	NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQTKNNMTIDGKELTISPA	300
QY	301	YLLWDLAISOSKQDEISASRPEDNEELRYSLRSTERHAPWVRNPIFVINGQIPSWLNL	360
DB	301	YLLWDLAISOSKQDEISASRPEDNEELRYSLRSTERHAPWVRNPIFVINGQIPSWLNL	360
QY	361	DNPRVTIVTHQDVFNLSHLPSTSSPAIESHIHRIEGLSKFYLLNDDVMFGKDVWPDF	420
DB	361	DNPRVTIVTHQDVFNLSHLPSTSSPAIESHIHRIEGLSKFYLLNDDVMFGKDVWPDF	420
QY	421	YSHSGKQKVLTPVPNCAECGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG	480
DB	421	YSHSGKQKVLTPVPNCAECGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG	480
QY	481	GGGTGSLGVGHPQFGGGINSVSCNQCANSWLADKFCDOACNVLSCGPDAGCCQDHF	540
DB	481	GGGTGSLGVGHPQFGGGINSVSCNQCANSWLADKFCDOACNVLSCGPDAGCCQDHF	540
QY	541	HELKVKILLNQTHIIPKGECLPYFSGFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL	600
DB	541	HELKVKILLNQTHIIPKGECLPYFSGFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL	600
QY	601	IMHSGMATIHFNLITQNTNDEEFKQITVEVDTRGPKLNSTAGKYNLVSPTILP	660
DB	601	IMHSGMATIHFNLITQNTNDEEFKQITVEVDTRGPKLNSTAGKYNLVSPTILP	660
QY	661	EAEILFEDIPEKRPFKRHDVNSTRAQAEVKIPLVNLISLLPKDAQLSLNTLDLQLEH	720

DB	661	EAEILFEDIPEKRPFKRHDVNSTRAQAEVKIPLVNLISLLPKDAQLSLNTLDLQLEH	720
QY	721	GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVAPOEKQVHKSLPNSLG	780
DB	721	GDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVAPOEKQVHKSLPNSLG	780
QY	781	VSELRQRLTTPAVSVKVNHGHDGQGNPDLDETTARFRVETHQTKTIGNVTKEPPSLIV	840
DB	781	VSELRQRLTTPAVSVKVNHGHDGQGNPDLDETTARFRVETHQTKTIGNVTKEPPSLIV	840
QY	841	PLESOMTKKTKITGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYF	900
DB	841	PLESOMTKKTKITGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYF	900
QY	901	QDLDEESLKTQLAYFTDSKNTGRQLK	928
DB	901	QDLDEESLKTQLAYFTDSKNTGRQLK	928

RESULT 9

US-10-023-888-2
 ; Sequence 2, Application US/10023888
 ; Publication No. US20030119088A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, William
 ; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
 ; FILE REFERENCE: 203515US77
 ; CURRENT APPLICATION NUMBER: US/10/023,888
 ; CURRENT FILING DATE: 2001-12-21
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 1199
 ; TYPE: PRT
 ; ORGANISM: hybrid
 US-10-023-888-2

Query Match	94.4%;	Score 4630;	DB 14;	Length 1199;
Best Local Similarity	99.3%;	Pred. No. 0;		
Matches	875;	Conservative 2;	Mismatches 4;	Indels 0; Gaps 0;
QY	45	SRDQYHVLFDSDYRNIAGKSFQRLCLPMPIDVYTWNGTDLLEKKELOQVREOEEQ	104	
DB	35	SRDQYHVLFDSDYRNIAGKSFQRLCLPMPIDVYTWNGTDLLEKKELOQVREOEEQ	94	
QY	105	KAMREILGNKTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSPSYFHSAS	164	
DB	95	KAMREILGNKTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSPSYFHSAS	154	
QY	165	SDIFNVAKPNPSTNVSVVVFSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLM	224	
DB	155	SDIFNVAKPNPSTNVSVVVFSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLM	214	
QY	225	QDLAFSLGPPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQ	284	
DB	215	QDLAFSLGPPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQ	274	
QY	285	KNNMTIDGKELTISPAYLLWDLAISOSKQDEISASRPEDNEELRYSLRSTERHAPWVR	344	
DB	275	KNNMTIDGKELTISPAYLLWDLAISOSKQDEISASRPEDNEELRYSLRSTERHAPWVR	334	
QY	345	NIFVINGQIPSWLNDNPRVTIVTHQDVFNLSHLPSTSSPAIESHIHRIEGLSKFYI	404	
DB	335	NIFVINGQIPSWLNDNPRVTIVTHQDVFNLSHLPSTSSPAIESHIHRIEGLSKFYI	394	
QY	405	LNDDVMFGKDVWPDFYSHSGKQKVLTPVPNCAECGPGSWIKDGYCDKACNNSACDWD	464	
DB	395	LNDDVMFGKDVWPDFYSHSGKQKVLTPVPNCAECGPGSWIKDGYCDKACNNSACDWD	454	
QY	465	GGDCSGNSGGGRIYAGGGTGSIGVGHWPQFGGGINSVSCNQCANSWLADKFCDOACN	524	
DB	455	GGDCSGNSGGGRIYAGGGTGSIGVGHWPQFGGGINSVSCNQCANSWLADKFCDOACN	514	

QY 525 VLSGFDAGDCGQDHFEHLYKVIILLPNQTHYIIIPKGECLPYFSAEVAKRGVEGAYSDNP 584
DB 515 VLSGFDAGDCGQDHFEHLYKVIILLPNQTHYIIIPKGECLPYFSAEVAKRGVEGAYSDNP 574
QY 585 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTQNTDEEFKMQITVEVDTRREGPKLNST 644
DB 575 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTQNTDEEFKMQITVEVDTRREGPKLNST 634
QY 645 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 704
DB 635 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 694
QY 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 764
DB 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 754
QY 765 PQEKQVHKSIILPNSLGVSRERLQRTFFPAVSVKVNGHDQGNPPLDLETTARFRVETHQK 824
DB 755 PQEKQVHKSIILPNSLGVSRERLQRTFFPAVSVKVNGHDQGNPPLDLETTARFRVETHQK 814
QY 825 TIGGNVTEKPPSLIVPLESQMTKEKKTIGKEKNSRMEENAHNIGVTEVLLGRKLQHY 884
DB 815 TIGGNVTEKPPSLIVPLESQMTKEKKTIGKEKNSRMEENAHNIGVTEVLLGRKLQHY 874
QY 885 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNTR 925
DB 875 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNTR 915

RESULT 10

US-10-023-889-2
; Sequence 2, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH
; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 2035120S77
; CURRENT APPLICATION NUMBER: US/10/023.889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid

US-10-023-889-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 45 SRDQVHVLFDSDYRDNIAKSFQNRCLPMPIDVVYTWNGTDLLELLKELQOVRQMBEEQ 104
DB 35 SRDQVHVLFDSDYRDNIAKSFQNRCLPMPIDVVYTWNGTDLLELLKELQOVRQMBEEQ 94
QY 105 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 164
DB 95 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 154
QY 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVMRGYLTDTKEVPGVLVM 224
DB 155 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVMRGYLTDTKEVPGVLVM 214
QY 225 QDLAPLSGFPPTFKETNOLTKLPENLSKVKLLQLYSEASVALLKLNPKDFOLNKQT 284
DB 215 QDLAPLSGFPPTFKETNOLTKLPENLSKVKLLQLYSEASVALLKLNPKDFOLNKQT 274
QY 285 KKNMTIDGKELTISPAYLLWDLISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWR 344
DB 275 KKNMTIDGKELTISPAYLLWDLISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWR 334

QY 345 NIFIVTNGOIPSWNLNDNPRVTIVTHQDVFRNLSHLPFPSSPAIESHIHRIEGLSOKFIY 404
DB 335 NIFIVTNGOIPSWNLNDNPRVTIVTHQDVFRNLSHLPFPSSPAIESHIHRIEGLSOKFIY 394
QY 405 LNDVDMFGKDVMPDDPYSHSGKQKYYLTWPVPNCABEGCPGSGWIKDGYCDKACNNSACDWD 464
DB 395 LNDVDMFGKDVMPDDPYSHSGKQKYYLTWPVPNCABEGCPGSGWIKDGYCDKACNNSACDWD 454
QY 465 GDCSCNSGGSYIAGGGGTGIGVCHPMQFGGGINSVYCNQGCANSWLADKFCDDQACN 524
DB 455 GDCSCNSGGSYIAGGGGTGIGVCHPMQFGGGINSVYCNQGCANSWLADKFCDDQACN 514
QY 525 VLSGFDAGDCGQDHFEHLYKVIILLPNQTHYIIIPKGECLPYFSAEVAKRGVEGAYSDNP 584
DB 515 VLSGFDAGDCGQDHFEHLYKVIILLPNQTHYIIIPKGECLPYFSAEVAKRGVEGAYSDNP 574
QY 585 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTQNTDEEFKMQITVEVDTRREGPKLNST 644
DB 575 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTQNTDEEFKMQITVEVDTRREGPKLNST 634
QY 645 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 704
DB 635 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 694
QY 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 764
DB 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 754
QY 765 PQEKQVHKSIILPNSLGVSRERLQRTFFPAVSVKVNGHDQGNPPLDLETTARFRVETHQK 824
DB 755 PQEKQVHKSIILPNSLGVSRERLQRTFFPAVSVKVNGHDQGNPPLDLETTARFRVETHQK 814
QY 825 TIGGNVTEKPPSLIVPLESQMTKEKKTIGKEKNSRMEENAHNIGVTEVLLGRKLQHY 884
DB 815 TIGGNVTEKPPSLIVPLESQMTKEKKTIGKEKNSRMEENAHNIGVTEVLLGRKLQHY 874
QY 885 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNTR 925
DB 875 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNTR 915

RESULT 11

US-10-023-890-2
; Sequence 2, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH
; TITLE OF INVENTION: MAMMALIAN CELLS
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023.890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid

US-10-023-890-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 45 SRDQVHVLFDSDYRDNIAKSFQNRCLPMPIDVVYTWNGTDLLELLKELQOVRQMBEEQ 104
DB 35 SRDQVHVLFDSDYRDNIAKSFQNRCLPMPIDVVYTWNGTDLLELLKELQOVRQMBEEQ 94
QY 105 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 164
DB 95 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 154
QY 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVMRGYLTDTKEVPGVLVM 224

Db	155	SDIFNVAKPNPSTNVS	VVFDSTKVEDAHSGLLKGNRSQTVMRGYLTATDKVPGVLVM	214
QY	225	QDLAFLSGFPPTFKETNQLTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKT		284
Db	215	QDLAFLSGFPPTFKETNQLTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKT		274
QY	285	KKNMTIDCKELTISPAYLLWDLSSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVR		344
Db	275	KKNMTIDCKELTISPAYLLWDLSSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVR		334
QY	345	NIFIVTNGQIPSWNLNDNPRVTIIVTHQDVFNRLSHLPTFSSPAIESHRIEGLSQFIY		404
Db	335	NIFIVTNGQIPSWNLNDNPRVTIIVTHQDVFNRLSHLPTFSSPAIESHRIEGLSQFIY		394
QY	405	LNDVFMFGKOWPDDFYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD		464
Db	395	LNDVFMFGKOWPDDFYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD		454
QY	465	GGDCSGNSGSGRYIAGGGGTGSGVGHWPQFGGINSVSYCNQGCANSLADKFCDOACN		524
Db	455	GGDCSGNSGSGRYIAGGGGTGSGVGHWPQFGGINSVSYCNQGCANSLADKFCDOACN		514
QY	525	VLSCGFDAGDCQDHFHLYKVILLPNQTHYIIPKGECLPYFSGFAEVAKEGVEGAYSDNP		584
Db	515	VLSCGFDAGDCQDHFHLYKVILLPNQTHYIIPKGECLPYFSGFAEVAKEGVEGAYSDNP		574
QY	585	IIRHASIANKWKTTHLIMHSGMNATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNST		644
Db	575	IIRHASIANKWKTTHLIMHSGMNATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNST		634
QY	645	AQKGYENLVSPITLLPEAILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLIP		704
Db	635	AQKGYENLVSPITLLPEAILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLIP		694
QY	705	KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA		764
Db	695	KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA		754
QY	765	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHDQGNPPDLLETTARFVETHQK		824
Db	755	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHDQGNPPDLLETTARFVETHQK		814
QY	825	TIGGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY		884
Db	815	TIGGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY		874
QY	885	TDSYLGFLPWEKKYFODLDEESLKTQLAYFTDSKNTR		925
Db	875	TDSYLGFLPWEKKYFODLDEESLKTQLAYFTDSKNTR		915

RESULT 12
US-10-024-197-2
; Sequence 2, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBSIDASE AND METHODS
; TITLE OF INVENTION: TREATING GAUCHER'S DISEASE
; FILE REFERENCE: 209794US0

RESULT 12

US-10-024-197-2

; Sequence 2, Application US/10024197

; Publication No. US20030133924A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBOSIDASE

; FILE REFERENCE: 209794US0

; CURRENT APPLICATION NUMBER: US/10/024,197

; CURRENT FILING DATE: 2001-12-21

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 2

; LENGTH: 1199

; TYPE: PRT

; ORGANISM: hybrid

US-10-024-197-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;

Best Local Similarity 99.3%; Pred. No. 0;

Qy	45	SRDQYHVLFDSDYRDNIAKGSFQNLCLPMPIDVVYVTVWNGTDLLELLKELQOVREQMBEEQ	104
Db	35	SRDQYHVLFDSDYRDNIAKGSFQNLCLPMPIDVVYVTVWNGTDLLELLKELQOVREQMBEEQ	94
Qy	105	KAMREILGNVTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPSHSA	164
Db	95	KAMREILGNVTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPSHSA	154
Qy	165	SDIFNVAKPNPSTNVS	224
Db	155	SDIFNVAKPNPSTNVS	214
Qy	225	QDLAFLSGFPPTFKETNQLTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKT	284
Db	215	QDLAFLSGFPPTFKETNQLTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKT	274
Qy	285	KKNMTIDCKELTISPAYLLWDLSSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVR	344
Db	275	KKNMTIDCKELTISPAYLLWDLSSAISOSKODEDISASRFEDNEELRYSLRSIERHAPWVR	334
Qy	345	NIFIVTNGQIPSWNLNDNPRVTIIVTHODVFNRLSHLPTFSSPAIESHRIEGLSQFIY	404
Db	335	NIFIVTNGQIPSWNLNDNPRVTIIVTHODVFNRLSHLPTFSSPAIESHRIEGLSQFIY	394
Qy	405	LNDVFMFGKOWPDDFYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD	464
Db	395	LNDVFMFGKOWPDDFYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD	454
Qy	465	GGDCSGNSGSGRYTAGGGGTGSGVGHWPQFGGINSVSYCNQGCANSLADKFCDOACN	524
Db	455	GGDCSGNSGSGRYTAGGGGTGSGVGHWPQFGGINSVSYCNQGCANSLADKFCDOACN	514
Qy	525	VLSCGFDAGDCQDHFHLYKVILLPNQTHYIIPKGECLPYFSGFAEVAKEGVEGAYSDNP	584
Db	515	VLSCGFDAGDCQDHFHLYKVILLPNQTHYIIPKGECLPYFSGFAEVAKEGVEGAYSDNP	574
Qy	585	IIRHASIANKWKTTHLIMHSGMNATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNST	644
Db	575	IIRHASIANKWKTTHLIMHSGMNATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNST	634
Qy	645	AQKGYENLVSPITLLPEAILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLIP	704
Db	635	AQKGYENLVSPITLLPEAILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNISSLIP	694
Qy	705	KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA	764
Db	695	KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA	754
Qy	765	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHDQGNPPDLLETTARFVETHQK	824
Db	755	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHDQGNPPDLLETTARFVETHQK	814
Qy	825	TIGGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY	884
Db	815	TIGGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHY	874
Qy	885	TDSYLGFLPWEKKYFODLDEESLKTQLAYFTDSKNTR	925
Db	875	TDSYLGFLPWEKKYFODLDEESLKTQLAYFTDSKNTR	915

RESULT 13

US-10-023-894-2

; Sequence 2, Application US/10023894

; Publication No. US20030143669A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; APPLICANT: KORNFIELD, Stuart

; TITLE OF INVENTION: EXPRESSION OF LYCOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-

; TITLE OF INVENTION: ACETYLGUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDAS

; FILE REFERENCE: 217139US77

; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid
US-10-023-894-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY	45	SRDQVHVLFDSDYRNIAGKSFQNLCLPMPIDVVTWNGTDLLELLELQVREOMEEQ	104
Db	35	SRDQVHVLFDSDYRNIAGKSFQNLCLPMPIDVVTWNGTDLLELLELQVREOMEEQ	94
QY	105	KAMREILGKNTTEPTKSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSPSPHSA	164
Db	95	KAMREILGKNTTEPTKSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSPSPHSA	154
QY	165	SDIFNVAKPKNPSTNVSVVWFDSTKDVDEDAHSGLLKGNRSQTWRGYLTTDKVPGVLVM	224
Db	155	SDIFNVAKPKNPSTNVSVVWFDSTKDVDEDAHSGLLKGNRSQTWRGYLTTDKVPGVLVM	214
QY	225	QDLAFLSGFPPTFKETNOLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKT	284
Db	215	QDLAFLSGFPPTFKETNOLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKT	274
QY	285	KNNMTIDGKELTISPAYLLDLSAISQSKODEDASRFDNEELRYSLRSIERHAPVVR	344
Db	275	KNNMTIDGKELTISPAYLLDLSAISQSKODEDASRFDNEELRYSLRSIERHAPVVR	334
QY	345	NIFVTNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPTSSPAESHIEGLSQFIY	404
Db	335	NIFVTNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPTSSPAESHIEGLSQFIY	394
QY	405	LNDVMFGKDWPDYFSSHSGQKVYLTWPVPCNACGPGSWIKDGYCDKACNNSACDW	464
Db	395	LNDVMFGKDWPDYFSSHSGQKVYLTWPVPCNACGPGSWIKDGYCDKACNNSACDW	454
QY	465	GGDCSGNSGGSRYPAGGGTSGISGVHPWQGGINSVSCNQCANSWLADKFCDOACN	524
Db	455	GGDCSGNSGGSRYPAGGGTSGISGVHPWQGGINSVSCNQCANSWLADKFCDOACN	514
QY	525	VLSGCFDAGDCGQDHFHELYKVILLPNQTHVIIIPKGECLPYFSPAFAVAKRGVAGSDNP	584
Db	515	VLSGCFDAGDCGQDHFHELYKVILLPNQTHVIIIPKGECLPYFSPAFAVAKRGVAGSDNP	574
QY	585	IIRHASIANKWKTTHLIMHSGMNAITTHFNLTFTONTNDEBFKQITVEVDTRGPKLNT	644
Db	575	IIRHASIANKWKTTHLIMHSGMNAITTHFNLTFTONTNDEBFKQITVEVDTRGPKLNT	634
QY	645	AQKGVENLVSPITLLPAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNTSLLP	704
Db	635	AQKGVENLVSPITLLPAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNTSLLP	694
QY	705	KDQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIIITDNTDLSVA	764
Db	695	KDQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIIITDNTDLSVA	754
QY	765	POEKQVHKSILPNSLGVSELRQRTFFAVSVKNGHDOGQNPPLDLETTARFVETHQK	824
Db	755	POEKQVHKSILPNSLGVSELRQRTFFAVSVKNGHDOGQNPPLDLETTARFVETHQK	814
QY	825	TIGGNVTKKEPPSLIVPLESOMTEKKITGKEKNSRMEENAHGIVTEVLLGRKLOHY	884
Db	815	TIGGNVTKKEPPSLIVPLESOMTEKKITGKEKNSRMEENAHGIVTEVLLGRKLOHY	874
QY	885	TDSYLGFLPWEKKYKFOQLDDEBSLTKQLAYFTDSKNTGR	925
Db	875	TDSYLGFLPWEKKYKFOQLDDEBSLTKQLAYFTDSKNTGR	915

RESULT 14
US-10-120-801-88
; Sequence 88, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Salomit
; APPLICANT: Smithson, Glennada
; APPLICANT: Gunther, Erik
; APPLICANT: Komaves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 88
; LENGTH: 847
; TYPE: PRT
; ORGANISM: human
US-10-120-801-88

Query Match 91.2%; Score 4473; DB 14; Length 847;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI	60
Db	1	MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI	60
QY	61	AGKSTQNRCLPMPIDVVTWNGTDLLELLELQVREOMEEQKAMREILGKNTTEPTK	120
Db	61	AGKSTQNRCLPMPIDVVTWNGTDLLELLELQVREOMEEQKAMREILGKNTTEPTK	120
QY	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSAADIENFAKPKNPSTNV	180
Db	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSAADIENFAKPKNPSTNV	180
QY	181	SVVWFDSTKDVDEDAHSGLLKGNRSQTWRGYLTTDKVPGVLVMDLAFSLGFPPTFKET	240
Db	181	SVVWFDSTKDVDEDAHSGLLKGNRSQTWRGYLTTDKVPGVLVMDLAFSLGFPPTFKET	240
QY	241	NGLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKTGKNTIDCKELTISPA	300
Db	241	NGLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKTGKNTIDCKELTISPA	300
QY	301	YLLMDLSAISQSKODEDASRFDNEELRYSLRSIERHAPVVRNIFVTNGQIPSWNLN	360

Db 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLSHLPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDWPPDF 420
Db 361 DNPRVTIVTHQDVFRNLSHLPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDWPPDF 420
Qy 421 YSHSGQKQVLTWPVPCABGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480
Db 421 YSHSGQKQVLTWPVPCABGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480
Qy 481 GGGTSGIVGHPWQFGGGINSVSCNQCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Db 481 GGGTSGIVGHPWQFGGGINSVSCNQCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNATTIHFNLTFQNTNDEEFKQIITVEVDTRGPKLNSTAKQYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTFQNTNDEEFKQIITVEVDTRGPKLNSTAKQYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKRPFPKRDVNSTRAQEEVKIPLVNISSLPKDAQSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKRPFPKRDVNSTRAQEEVKIPLVNISSLPKDAQSLNTLDLQLEH 720
Qy 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
Db 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
Qy 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840
Db 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840
Qy 841 PLESQM 846
Db 841 PLESQM 846

RESULT 15

US-10-094-749-2392
; Sequence 2392, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIC, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2392

; LENGTH: 846
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2392

Query Match 91.1%; Score 4468; DB 14; Length 846;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLFKLLQROTTCLSHRYGLVVCFLGVVTTVSFAQFGEVLEWSRDQYHVLFDSDYRDN 60
Db 1 MLFKLLQROTTCLSHRYGLVVCFLGVVTTVSFAQFGEVLEWSRDQYHVLFDSDYRDN 60
Qy 61 AGKSFQNRCLCPMPIDVVVYTWVNGTDLLELLKELQVREQMEEEKAMREILGNKTTEPTK 120
Db 61 AGKSFQNRCLCPMPIDVVVYTWVNGTDLLELLKELQVREQMEEEKAMREILGNKTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Qy 181 SVVVFDS TKDVEDAHSGLLKGNRSQTVMRGYLTITDKEVPGLVLMQDLAFSLGPPPTFKET 240
Db 181 SVVVFDS TKDVEDAHSGLLKGNRSQTVMRGYLTITDKEVPGLVLMQDLAFSLGPPPTFKET 240
Qy 241 NQKTKLPENLSSKVKLLQLLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300
Qy 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPWVRNIFIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLSHLPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDWPPDF 420
Db 361 DNPRVTIVTHQDVFRNLSHLPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDWPPDF 420
Qy 421 YSHSGQKQVLTWPVPCABGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480
Db 421 YSHSGQKQVLTWPVPCABGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480
Qy 481 GGGTSGIVGHPWQFGGGINSVSCNQCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Db 481 GGGTSGIVGHPWQFGGGINSVSCNQCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNATTIHFNLTFQNTNDEEFKQIITVEVDTRGPKLNSTAKQYENLVSPITLLP 660
Db 601 IMHSGMNATTIHFNLTFQNTNDEEFKQIITVEVDTRGPKLNSTAKQYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKRPFPKRDVNSTRAQEEVKIPLVNISSLPKDAQSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKRPFPKRDVNSTRAQEEVKIPLVNISSLPKDAQSLNTLDLQLEH 720
Qy 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
Db 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSVAPOEKQVHKSILPNSLG 780
Qy 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840
Db 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840
Qy 841 PLESQM 846
Db 841 PLESQM 846

Search completed: November 21, 2004, 13:22:34
Job time : 104.171 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:39:03 ; Search time 28.3503 Seconds
(without alignments)
3149.495 Million cell updates/sec

Title: US-10-023-888-4
Perfect score: 4907
Sequence: 1 MLFKLQRTVTCLSHRYGL.....SLKTQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: Pirl1.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	734.5	15.0	384	2 I49528	hypothetical prote
2	322.5	6.6	586	2 T34703	hypothetical prote
3	314	6.4	541	2 T34701	hypothetical prote
4	266	5.4	545	2 C82014	probable capsule b
5	256	5.2	532	2 G70536	probable cpay prot
6	255.5	5.2	542	2 D87185	probable UDP-gluc
7	228.5	4.7	373	2 S42434	hypothetical prote
8	175.5	3.6	2352	2 T30201	Notch homolog prot
9	168.5	3.4	2471	2 A49128	cell-fate determin
10	166.5	3.4	1429	2 S06434	homeotic protein 1
11	163	3.4	2531	2 S18188	Notch protein homo
12	162.5	3.3	2437	2 A46019	Notch-1 protein -
13	162.5	3.3	2437	2 A42612	transmembrane prot
14	154.5	3.1	2321	2 S78549	Notch3 protein - h
15	152	3.1	1964	2 T09059	Notch4 - mouse
16	148	3.0	861	2 A48825	Notch homolog Motc
17	147	3.0	1203	2 A49175	Notch B protein -
18	147	3.0	2318	2 S45306	Notch 3 protein -
19	146.5	3.0	387	2 A49175	Notch A protein -
20	146	3.0	2531	2 T31070	Notch homolog - se
21	145.5	3.0	2555	2 A40043	Notch protein homo
22	144.5	2.9	1755	2 S70298	TYB protein - years
23	140.5	2.9	1755	2 S50663	TYB protein - years
24	138	2.8	2524	2 A35844	Notch protein - Af
25	137.5	2.8	1755	2 S69866	TYB protein - years
26	136.5	2.8	1701	2 A26868	major merozoite su
27	135.5	2.8	1701	2 A54498	major merozoite su
28	134	2.7	1295	2 A32901	glpI protein precu
29	133.5	2.7	540	2 T34702	Hypothetical prote

30	133.5	2.7	1969	2 T08875	histidine kinase h
31	133.5	2.7	2052	2 C97038	phage-related prot
32	131.5	2.7	957	2 C69463	type I restriction
33	131.5	2.7	1422	2 T18404	chromatin remodel
34	131.5	2.7	1640	2 A24594	probable major sur
35	130	2.6	484	2 T16595	hypothetical prote
36	129.5	2.6	1639	2 S05603	major merozoite su
37	129	2.6	1276	2 T09204	probable tail-host
38	129	2.6	2703	1 A24420	Notch protein - fr
39	128.5	2.6	756	2 C64236	hypothetical prote
40	128	2.6	644	2 T47835	DNA mismatch repa
41	127.5	2.6	780	2 B70112	TYB protein - years
42	127.5	2.6	1755	2 S69963	hypothetical prote
43	127	2.6	1459	2 S50437	TYB protein - years
44	126.5	2.6	1328	2 S52894	protein kinesin F2
45	125.5	2.6	987	2 B96766	

ALIGNMENTS

RESULT 1
I49528
hypothetical protein - mouse (fragment)
N;Alternate names: cDNA1
C;Species: Mus musculus (house mouse)
C;Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004
C;Accession: I49528
R;Cordes, S.P.; Barsh, G.S.
Cell 79, 1025-1034, 1994
A;Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper tr
A;Reference number: A55200; MUID:95094266; PMID:8001130
A;Accession: I49528
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-384 <RES>
A;Cross-references: UNIPROT:Q61340; GB:I36434; NID:9625041; PIDN:AAA65688.1; PID:96250

Query Match 15.0%; Score 734.5; DB 2; Length 384;
Best Local Similarity 57.9%; Pred. No. 2.1e-38;
Matches 165; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

Qy	644	TQKGVENAVSPITLLPEREILFEDIPKEKPKFKRDVNSTRAQBEVKIPLVNISLL	703
Db	3	TTQKAYESLVSVPTLPQADVFEDVPKEKPKFKIRHDVNAIGFQBEVKIPLVNISLL	62
Qy	704	PKDAQSLNLTDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDSLV	763
Db	63	PXEAQVRLSNLDLQLEHGDITLKGYNLSKALLRSFLGNSLDTKIKPQA-RDDETKGNLE	121
Qy	764	APQEQVHKSILPNSLGVSRQLRITFPFPAVSVKNGHOGQNPPLDLETTFARFVETHTQ	823
Db	122	VPQENPSHRR--PHGFACEHRSERWTAFATVTVKGRDHALNPPFPVLETNARL----	174
Qy	824	KTIGGNVTKEKPPSLIVPLESQMTKEKITGKEKNSRMEENAHNIGVTEVLLGRKLQH	883
Db	175	PILGVTVSKENSLIVVPESHLP-----KEESDRAEGNA---VPVKELVPORCSR	224
Qy	884	YTDSYLGFLPWKKYFQDLDERESLTKQLAYFTDSKNTGRQLK	928
Db	225	IIQA---FCPGKKKYFQDLDAEESLTKQLAYFTDSKNTGRQLK	266

RESULT 2
T34703
Hypothetical protein SC1C3.11 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C;Accession: T34703
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21554
A;Accession: T34703

A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-586 <OLI>
A:Cross-references: UNIPROT:O69853; EMBL:AL023702; PIDN:CAA19235.1; GSPDB:GN00070; SCOE
C:Genetics: strain A3(2)
A:Gene: SCOE:SC1C3.11

Query Match 6.6%; Score 322.5; DB 2; Length 586;
Best Local Similarity 46.3%; Pred. No. 2.5e-12;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;

QY 290 IDGKELTISPAYLLW--DLAISQSKQDEIDISARFEDNEELYSLSRTERHAPVVRNIF 347
DB 282 VDGND---PA---WRRRRSAYDGGYHAESANAARVYSRDELYSURALEQNAQWVRHVH 334
QY 348 IVTNGQIPSWLNDNPRVTIVTHQVFNRLSHLPFTSSPAIESHRIEGLSQKFIYLLND 407
DB 335 LVTDGQPAWLNDSHPALTVDHSEIFADPAALPTFNSHAIESRLHHIKGLSEHFLYLLND 394
QY 408 DVMPGKDWPPDDFY 421
DB 395 DMFLGRPVTQDDF 408

RESULT 3
T34701
hypothetical protein SC1C3.09 SC1C3.09 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T34701
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21554
A:Accession: T34701
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-541 <OLI>
A:Cross-references: UNIPROT:O69851; EMBL:AL023702; PIDN:CAA19233.1; GSPDB:GN00070; SCOE
C:Genetics: strain A3(2)
A:Gene: SCOE:SC1C3.09

Query Match 6.4%; Score 314.; DB 2; Length 541;
Best Local Similarity 39.2%; Pred. No. 7.5e-12;
Matches 67; Conservative 31; Mismatches 63; Indels 10; Gaps 5;

QY 290 IDGKELTISPAYLLWLSAISQSKQDEIDISARFEDNEELYSLSRTERHAPVVRNIF 349
DB 242 VDGND---PAWQKQAKAGEVYHAESASDARFISRDELYSIRSLHLPFAPVIRNIYV 297
QY 350 TNGQIPSWLNDNPRVTIVTHQVFNRLSHLPFTSSPAIESHRIEGLSQKFIYLLNDV 409
DB 298 TDDQVPAWRDLPCARIATHREIFRPEDLPFTNSHIESQLHHIEGLAEHFLYNDMM 357
QY 410 MFGKDWPPDDFYSHSKQKYLTPVPN-CAGCGPGSWIKDGYCDKACNNS 459
DB 358 FMGRFVAPHSFT-PNGTARY--FPRSRNPQGAFAE--TDSVPDAACKGN 403

RESULT 4
C82014
probable capsule biosynthesis protein NMA0200 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C:Accession: C82014
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: C82014
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-545 <PAR>
A:Cross-references: UNIPROT:Q9JWW8; GB:AL157959; NID:g378776; PIDN:CA88351
C:Genetics: serogroup A, strain Z2491
A:Gene: sacB; NMA0200

Query Match 5.4%; Score 266.; DB 2; Length 545;
Best Local Similarity 27.8%; Pred. No. 7.6e-09;
Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;

QY 145 LPANITLKVPSLYPSFHSASDIFNVAKPKNPSTNVSVVDFSTKVDHAHGLKX-NS 203
DB 97 LPSNLTLLK--PALCILESHKEDFLN-----KFLLTISSEMLKQYKFGNGIKNPKS 145
QY 204 RQTVRGY-----LTTQKEVPGLV-----LMQDLAFLSCFPPTFKETNQLKT 245
DB 146 VNEIWTDLFSIAHVDMLKSTDRILSSISQFWFLEFKEDKQFI-----LFPTANRYSR 200
QY 246 KLPELSSKVKLLQLYSEASVALLKANNPKDFQSLNPKQTKNWTIDGKELTISPAYLLW 305
DB 201 KLRH-----SIKNQLFKEG-----IRNYSEISLPEEDHNPDI DLVFTWNSDKNWQ 251
QY 306 LSASISQSKQD---EDISARFEDNEELYSLSRTERHAPVVRNIFIVTNGQIPSWLNDN 362
DB 252 -ELYKVKPDPFNSDATSTSRFLSRDELKALRSWMNGSFIRKIFIVSNCAAPPALWDLNN 310
QY 363 PRVTIVTHQVFNRLSHLPFTSSPAIESHRIEGLSQKFIYLLNDVDFGKDWPPDDFY 422
DB 311 PKIQWVHEEIMPQ-SALPTFSSHAIEISLHIFGINSYFIYSNDDFLLLKPLNKDNFFY 369
QY 423 HSKGQKQVYL-TWPVPN--CAEGCP 443
DB 370 SNGIAKLRLAENGNGVNGECTEGEP 393

RESULT 5
G70536
probable cpsY protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: G70536
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70536
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-532 <COL>
A:Cross-references: UNIPROT:O06628; GB:Z95618; GB:AL123456; NID:g3261788; PIDN:CA809105
C:Genetics: strain H37RV
A:Gene: cpsY

Query Match 5.2%; Score 256.; DB 2; Length 532;
Best Local Similarity 39.6%; Pred. No. 3.1e-08;
Matches 53; Conservative 30; Mismatches 45; Indels 6; Gaps 3;

QY 290 IDGKELTISPAYLLWLSAISQSKQDEIDISARFEDNEELYSLSRTERHAPVVRNIF 348
DB 210 VDGSD---PBFRRARMQMSQVYVGECDAAIRQIDELKYAIRSVNMFAPVIRIFI 265
QY 349 VTNGQIPSWLNDNPRVTIVTHQVFNRLSHLPFTSSPAIESHRIEGLSQKFIYLLND 408
DB 266 ATDSFTPPWL-AEHPKTIIVRAEDHFSRSLPTVNSHAVESQLHHIPGLSEHFLYNSDD 324
QY 409 VMFGKDWPPDDFY 422
DB 325 MFFGRPLKASMPFS 338

Qy	271	LNPNKDFQELNKTQKNMTDGHKELTSPAYLLWD-----LSAISQSKQDEDI-----S 319
Db	43	INADANLASITENVLPNFPIDVVFTVDNTDKAQEQYRTLPIDQ-----EDIGLXATD 98
Qy	320	ASRFEDNEELRYSIRIERHAPVVRNFIIVTNGQIPSWNLND-NPRVTIVTHQDVFRNLS 378
Db	99	PARFSNHNELFYSAQAKFQWPMVVRNFIIVTADQPKWLDENTHSKIKLINHSQLI-DAK 157
Qy	379	HLTFSPATESIHRIEGHSQKFIYNDNDMFQKDVWPDFFVSHS 424
Db	158	YLPFTSHVIEANLYKLPIDLSEHFIFYFNDDVFVARPLMPNHFENN 203
RESULT 8		
T3201		
Notch homolog protein - sea squirt (Haloecynthia roretzi)		
C;Species: Haloecynthia roretzi		
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 08-Sep-2002		
C;Accession: T320201		
R;Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.		
Dev. Genes Evol. 207, 371-380, 1997		
A;Title: Notch homologue from Haloecynthia roretzi is preferentially expressed i		
A;Reference number: Z20775		
A;Accession: T320201		
A;Status: Preliminary; translated from GB/EMBL/DBDJ		
A;Molecule type: mRNA		
A;Residues: 1-2352 <HOR>		
A;Cross-references: EMBL;AB001327; NID:dl204472; PID:dl026501; PIDN:BA25571.1		
C;Genetics:		
A;Gene: Notch		
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology		
Query Match 3.6%; Score 175.5; DB 2; Length 2352;		
Best Local Similarity 20.3%; Pred. No. 0.034;		
Matches 112; Conservative 70; Mismatches 190; Indels 179; Gaps 26;		
Qy	438	CABGCPGSWIK-DGYCDK--ACNNSACMDGGBDCS-GNSGSRYSIAGGGTGTGIVGHPW 493
Db	1308	CSNTC--ELVKGDIGCDKEICGVNHECFDGGDCSLGIS-----ILGGCKDKELRCWQLY 1360
Qy	494	QFGGGINSVSYCNQGC-----ANSWLADKFCDCACNVLS 527
Db	1361	QDG-----HCDEHNEDCLHDGMDCATPVGGCNPKVEKYCNEYNNDYCDGGCKNAN 1413
Qy	528	CGFDAGDCGDH---FHELYKVILLNQTHYIIPKGECLPYFSAEVAKEGVGEGAYSDNP 584
Db	1414	CGWDGADCIKDYPAQYARGFLVIRMEGITPEEMREQEMIFIEFGKILRSIE----- 1466
Qy	585	IIRHASIANKWKTIHLIMSGMATTIHNLTTCNT-----NDEEPKQIITVEVDRE-- 637
Db	1467	ITISIEI-----FLESSRRRALLDPDVMKSTRHRATDTSATSEVKVRLDNRLCL 1518
Qy	638	-----GPKLNSTAQGYENL---VSPITLLPEAILFEDIPKEKRPF----- 676
Db	1519	LNDDSCFWBASAAASYVSASASTGLQVSPDVSTATSQP-----DETSDPETPLVYIL 1571
Qy	677	-----KPKHVDNSTRAQAEVVKPL-----VNISLLPK 705
Db	1572	SGSVGVLLVLLVAVILVRQKRKRENGTLMIPGEGVFPKKQKREPIQGDQDNLNKKLSK 1631
Qy	706	DAQLSNTLTDQLEHGDTILKGYNLKSALLRFLMNSCHAKIKQAIIITDETNDLSVAP 765
Db	1632	SSQLDENSSTPFIHSHWEDGNQKHQSKYLAQDESLSGSSLPVYTNQS---KFNLSAL-- 1684
Qy	766	QEQVHKSI---LPNSL---GYSERLQRTFFPAVSVKVNGHDQGNPPLDETATPRVE 819
Db	1685	SDSEAPDPVRKWTFOHQLQAGLSELL-----PSSPLTENGQDLND---IDI----- 1727
Qy	820	THTQKTIGGNVTKKPPSLIVPLESQMTKEK---KITCKEKNRMEENAEHNIG--VTEV 875
Db	1728	-----RCPDGHTPLMVASTRGGGIEVYSDSAGLSGEGSNI IANLLTO-1772

QY 876 LLGRKLOHYTD 886
 Db 1773 --GASLSAQTD 1781

RESULT 9
 A49128
 cell-fate determining gene Notch2 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 16-Aug-2004
 C:Accession: A49128
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 116, 931-941, 1992
 A:Title: Notch2: a second mammalian Notch gene.
 A:Reference number: A49128; MUID:93202015; PMID:1295745
 A:Accession: A49128
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2471 <WEI>
 A:Cross-references: UNIPROT:Q9QW30
 A:Experimental source: Schwann cell
 A:Note: sequence extracted from NCBI backbone (NCBIP:127811)
 C:Superfamily: Notch protein; ankyrin repeat homology; EGF homology
 F:264-295/Domain: EGF homology <EGX1>
 F:799-830/Domain: EGF homology <EGF1>
 F:877-908/Domain: EGF homology <EGX2>
 F:1029-1060/Domain: EGF homology <EGF>
 F:1067-1098/Domain: EGF homology <EGX3>
 F:1153-1184/Domain: EGF homology <EGF3>
 F:1191-1222/Domain: EGF homology <EGX4>
 F:1876-1908/Domain: ankyrin repeat homology <AN1>
 F:1909-1941/Domain: ankyrin repeat homology <AN2>
 F:1943-1975/Domain: ankyrin repeat homology <AN3>
 F:1976-2008/Domain: ankyrin repeat homology <AN4>
 F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 3.4%; Score 168.5; DB 2; Length 2471;
 Best Local Similarity 26.2%; Pred. No. 0.1;
 Matches 44; Conservative 15; Mismatches 58; Indels 51; Gaps 5;

QY 417 PDDFYSHKQKQVLTWPNVNCAGGCGSWIKDGYCDKACNNSACDWDGDCSGSGGR 476
 Db 1404 PPFMGSHCESYAPTSTPTATCLISQYCADKARDGICDEACNNSACQWDGDC----- 1455

QY 477 YIAGGGTSGIVGHPW-QFEGGINSVSYCNQGC-----AN 511
 Db 1456 -----SUTMEDPMANTSSLRWEYINQCDLNTAECLFDFECQNSKTCYD 1506

QY 512 SWLADKF-----CDQACNVLSGDFDAGCGQDHFHLYK-----VILLP 550
 Db 1507 KYCADHFKDNCXKGNNECGWGLDCAADQDENLAEGILVIVLLP 1554

RESULT 10
 S06434
 homeotic protein lin-12 precursor - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Aug-2004
 C:Accession: S06434; A24769
 R:Yochem, J.; Weston, K.; Greenwald, I.
 Nature 335, 547-550, 1988
 A:Title: The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with ove
 A:Reference number: S06434; MUID:88334747; PMID:3419531
 A:Accession: S06434
 A:Molecule type: DNA
 A:Residues: 1-1429 <YOC>
 A:Cross-references: UNIPROT:P14585; EMBL:M12069; NID:g156357; PIDN:AAA70191.1; PID:g1563
 R:Greenwald, I.
 Cell 43, 583-590, 1985
 A:Reference number: A24769; MUID:86079540; PMID:3000611
 A:Accession: A24769
 A:Molecule type: DNA
 A:Residues: 173-712 <GRE>

C:Genetics:
 A:Introns: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3
 C:Superfamily: ankyrin repeat homology; EGF homology
 C:Keywords: glycoprotein; transmembrane protein
 F:254-284/Domain: EGF homology <EGF1>
 F:507-540/Domain: EGF homology <EGF>
 F:547-578/Domain: EGF homology <EGF2>
 F:909-931/Domain: transmembrane #status predicted <TM>
 F:1093-1125/Domain: ankyrin repeat homology <AN1>
 F:1206-1238/Domain: ankyrin repeat homology <AN2>
 F:1240-1272/Domain: ankyrin repeat homology <AN3>

Query Match 3.4%; Score 166.5; DB 2; Length 1429;
 Best Local Similarity 20.4%; Pred. No. 0.058;
 Matches 113; Conservative 43; Mismatches 172; Indels 225; Gaps 22;

QY 438 CAEGCPGSMKDGCDKACNNSACDWDGDCSGNS---GGSRYTAGGGTSGIVGHPWQ 494
 Db 643 CSERA-----NDGNCADACNVACKFDGDCSGKREPFKCRY-----GNMCADF-- 687

QY 495 FGGINSVSYCNQGCAN-----SWLADKFCDQACNVLS 528
 Db 688 FANGV-----CNQACNNEELYDGMCLPAVVRCPVKIREHCASRFANGICDPECNTNGC 742

QY 529 GFDAGDCQDHFHLYKVVILLPNQTHYIIKGECLPYFSFAEVAKRGVEGAYSDNP1IRH 588
 Db 743 GFDGDC-----DNET----- 753

QY 589 ASIANKWKTHLIMHSGNNATTI-HFNLTQNTNDEEPK-----MOLITVEVD 634
 Db 754 -----NATIITNIRITVQ-MDPKSFQVTGGQSLMEISALEVTVRIQ 794

QY 635 -TREGPKLNSTAQGYENLVSPITL---LPEABILFEDPKERKPKRKHVDVNSTRA 689
 Db 795 RDEGLV---FQNGESEMVRVKNERQLTQHVLTSTISR-----KIKSATNIGVVV 846

QY 690 QEEVKIPLVNLISLLPKDAQSLNTLDLQLEHGDITLKGYNLS-----KSALLRSFL- 740
 Db 847 YLEQENCDTGKLYKDAQSVVDSISARLAKKIDSGFIPISALVAEPKSGNNTGFLS 906

QY 741 -----MNSOHAQKNOAIITDET 758
 Db 907 WNALLIAGCLIVVVLMLGALPCNTRKRMNASVWPPMENEKKNQHSITSSQ 966

QY 759 NDSLVAPOEKV-----HKSILPNSLVGYSE-----RLQRLTFPVAIVKVG 799
 Db 967 HSLLEASVDGYIKQRNELQHYSLVNPQGYGNGNDLFGDFNHTNLQIPTPEPEPIKL 1026

QY 800 HDQGNPPLDLETTARFVE---THQKT---TGNVTKKPPSLIVPLESQMTREKKI 852
 Db 1027 HTEAGSYAITEPITRESVNIIDPRNRTVLHWTASNSAEKSEDLIV---HEAKECIA 1082

QY 853 TGKEKENSMEEN 865
 Db 1083 AGADVNMDCDEN 1095

RESULT 11
 S18188
 notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S18188
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 113, 199-205, 1991
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383; PMID:1764995
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <WEI>
 A:Cross-references: EMBL:X57405; NID:57634; PID:G57635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>

```

Db      1596 VSHVLHTNWFKRDAQGQMIFFY:::-----GREELRKHPKRSV---GWATTSLL 1644
Qy      602 MHSG-----WNATTHFNLTFTQNDDEEFKMQITVEVDTR:::-----GP 639
Db      1645 PTNGRGHRELDPMWDHGSIVY:::-----LEIDNRQCVQSSSQCFQSATDVAA 1691
Qy      640 KLNSTAQGYENLYSPITLLPEAEILFEDIP-----KE 672
Db      1692 FLGALASLGSLNIPYKIEAV-KSETVEPLPSQLHLMYVAAAAFVLLFFVCGVLLSRKR 1750
Qy      673 KR-----FPK-FKRHDVNSTRRQAQEEVKIPL--VNISLLP-KDAQLSLNTLIDLQLSHG- 721
Db      1751 RRHQGLWFPFGFKVSEASKKRR-----PLGSDVGLKPLKNASDQALMDDNQNWGD 1805
Qy      722 -DITLKGYNLSKALLRFLMNSQHAKIKNQAITDETNDSLVAPQEKVHKSLPLNSLG 780
Db      1806 EDLETKKFRFEPVVLPLDQTDHRQWTOQHLDAADLRVSAVMAPTPPQ----- 1854
Qy      781 VSERLQRLTPFAVSKVNGHDQGNPPLDLETTARFVETHQTKTIGNVTKEKPPSLI 839
Db      1855 -----GEVDADCMQNVVRGPD--GFTFLMTASCSCGGLET-----GNSEEEEDAPAVI 1900

RESULT 12
A46019
notch-1 protein - mouse
N;Alternate names: notch protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text change 16-Aug-2004
C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
R;Gcl Amo, F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A;Reference number: A46019; MUID:93194170; PMID:8449489
A;Accession: A46019
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531 <DEL>
A;Cross-references: UNIPROT:Q01705; GB:Z111886; GB:S47228; NID:G288502; PIDN:CAA77941.1;
A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A;Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest
A;Reference number: S25144
A;Accession: S25144
A;Molecule type: mRNA
A;Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <PRA>

```

Db 1555 EWDGLDCA-EHVPRLAAGTLVLVLLPPDQLRNNSPHFL---RELSHVLHTNVVFKRDA 1610
 QY 577 EGAYSDNPIIRHASTANK-----WKTIIHLI-----MHSGMNTIHNLFQNT 620
 Db 1611 QQQQMFPPYHHEELRHPHKRSTVGWATSSLLPGTSGGRQRELPMDIRGSIVY--- 1667
 QY 621 NDPEFKQITVEVDTRE-----GPKLNTAOGYENLVSPITLLPEAEIL 665
 Db 1668 -----LEIDNRQCVOSSQCFQSATDVAFLGALASGLSNIPYKIEAV-KSEPV 1716
 QY 665 FEDIP-----KKEK-----FKP-KRHDVNSTRAQE 691
 Db 1717 EPLPSQLHMTVAFAAVFLVFFVCGVLLSRKRRRQHGQWFFPEGFKVSEASKKRR- 1775
 QY 692 EVKIFL--VNISLLP-KDAQLSLNTLDLQLEHG--DITLKGYNLSKALLRSLFMSQA 746
 Db 1776 ----PLGEDSVGLKPKVNASDCAALMDNDQNEWGDEDLTKKFRPEEPVLPDLSDQTDHR 1831
 QY 747 KIKNQAITDETNDLSVAPQEQVHKSLPNSLGVSELRQLTTPPAVSVKVNGHQGNP 806
 Db 1832 QWTOQHLDAADLRMGAMAPTTPQ-----GEVDADCMVNVVRGPD--GFT 1873
 QY 807 PLDETTFARFRVETHTKTIGNTVTEKPPSLI-----VPLESQMTKEKIT----- 853
 Db 1874 PLMIASCGGGLET-----GNSBEEDAPAVISDFIYQASLHNQTDRTGETALHAAR 1927
 QY 854 -CKEKENSMEENAEHNI-----GVTEVILL 877
 Db 1928 YSRDRKRLEASADANIQDNMGRTPLHAASVADAQGVFOILL 1970

RESULT 13

S42612
 transmembrane protein precursor - zebra fish
 C:Species: Brachydanio rerio (zebra fish)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
 C:Accession: S42612
 R:Bierkamp, C.; Campos-Ortega, J.A.
 Mech. Dev. 43, 87-100, 1993
 A:Title: A zebra fish homologue of the Drosophila neurogenic gene Notch and its pattern
 A:Reference number: S42612; MUID:94128602; PMID:8297791
 A:Accession: S42612
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2437 <BIE>
 A:Cross-references: UNIPROT:P46530; EMBL:X69088; NID:G433866; PIDN:CAA48831.1; PID:G4338
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:755-786/Domain: EGF homology <EGF1>
 F:1023-1054/Domain: EGF homology <EGF2>
 F:1185-1216/Domain: EGF homology <EGF3>
 F:1915-1947/Domain: ankyrin repeat homology <AN1>
 F:1948-1980/Domain: ankyrin repeat homology <AN2>
 F:1982-2014/Domain: ankyrin repeat homology <AN3>
 F:2015-2047/Domain: ankyrin repeat homology <AN4>
 F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 3.3%; Score 162.5; DB 2; Length 2437;
 Best Local Similarity 20.7%; Pred. No. 0.24;
 Matches 119; Conservative 58; Mismatches 201; Indels 197; Gaps 29;
 QY 421 YSHSKGQ-----KVLTFWPNCAEGCPGSGWIKDGYCDKXCNNSACDWDGDCGNS 472
 Db 1427 YSFGSGQGRDIAPPVEVEIRCEIAQC-EGRGNAI-----CDTQCNHACGWDGDC---- 1477
 QY 473 GGSRYAGGGGTGSGVGHPPQ-----FGGINSVSYCNQGCANS----- 512
 Db 1478 -----SLNFDDPNQNSAALQWRVFNDS-----KCDQCQATACGLYDGFPC 1519
 QY 513 -----WLADKFCQQAQCNVLSGCFDAGDCGQD-----HFHELYKVVILLP 550
 Db 1520 QRLSGQCNPYDQYCRDHYADGHGDCQGCNNAECWGLDCCADDVPQKLAVGSVLVWHIP 1579
 QY 551 -----NQTHYII-----PKGECL--PYFSFAEVAKRGVEGAYSDNPIIR 587

Db 1580 PDELNRSSSFLRELSSLLHTNVVFRDANGEALIFPYG-----SEHLSK 1626
 QY 588 HASTANKWKTIIHLMHSGMNTIHN-LTFQNTNDEEFKQKIT-VEVDTRE----- 637
 Db 1627 HK--RSDWTFDQGMQBARSLTSFLKPRTRRELDMHVKGSIVYLEIDNRQCFQSDRC 1684
 QY 638 -----GPKLNTAOGYENLVSPITLLPEABILPEDIKPKRFPKFRHVDVNSTRAQ 690
 Db 1685 FQSATDVAFLGALASSG--NLNV-----YILEAVTSEGPPK-----TG 1723
 QY 691 BEVKIPLVNLISLLPKDQLSLNTLDL-----OLEHGDITLKGYNLSKALLRSLFMSQ 744
 Db 1724 EYMFPLVLLAL-----AVLAAAGVVVSRKRREHGQWFF-----EGFKVNEP 1769
 QY 745 HAKIKNQAITDETNDLSVAPQEQVHKSLPNSL--GVSELRQLTTPPAVS-VKNG- 799
 Db 1770 KKKREP-----GEDSVGLKPKVNSDSSLMDSQLEMAEDDTNKFRFEGQSILEMSQ 1824
 QY 800 --HQGQNPPLDLETTARFRVETHTKTIGNVTK-----KPPSLIVPLESQMTKEKK 851
 Db 1825 LDHRQWTQOHL-----AADRLNSMAPTPQOQIENDCMVNVVRGPDGFTFL---MIASCS 1878
 QY 852 ITGKEKENSMEENAEHNI-GVTEVILLGRKLQHYTD 886
 Db 1879 GGGLENGEAGEEDPSADVTITDFIYHGANLHNQTD 1913

RESULT 14

S78549
 notch3 protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
 C:Accession: S78549; S71825
 R:Joutel, A.; Tournier-Lasserre, B.
 submitted to the EMBL Data Library, April 1997
 A:Reference number: S78549
 A:Accession: S78549
 A:Molecule type: mRNA
 A:Residues: 1-2321 <JOUT>
 A:Cross-references: UNIPROT:Q9UM47; EMBL:U97669; NID:G2668591; PIDN:AAB91371.1; PID:G266
 X:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabrier, H.; Mouton, P.; Alamowit
 X: M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, B.
 Nature 383, 707-710, 1996
 A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke
 A:Reference number: S71825; MUID:97032728; PMID:8878478
 A:Accession: S71825
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 67-113;138-194;268-333; 'G', 335-346;536-613;716-765;1240-1279;1815-1888 <JOU2
 A:Cross-references: EMBL:U97669
 C:Genetics:
 A:Gene: notch3
 A:Map position: 19p13.1
 C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 C:Keywords: tandem repeat; transmembrane protein
 F:123-155/Domain: EGF homology <EGX1>
 F:162-194/Domain: EGF homology <EGF1>
 F:240-271/Domain: EGF homology <EGX2>
 F:318-349/Domain: EGF homology <EGF>
 F:473-504/Domain: EGF homology <EGX3>
 F:853-884/Domain: EGF homology <EGF3>
 F:928-959/Domain: EGF homology <EGX4>
 F:1838-1870/Domain: ankyrin repeat homology <AN1>
 F:1871-1903/Domain: ankyrin repeat homology <AN2>
 F:1905-1937/Domain: ankyrin repeat homology <AN3>
 F:1938-1970/Domain: ankyrin repeat homology <AN4>
 F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match

Best Local Similarity 3.1%; Score 154.5; DB 2; Length 2321;
 Pred. No. 0.69;


```
Matches 38; Conservative 14; Mismatches 37; Indels 65; Gaps 6;
QY 433 WVPVNC-----AEGCPGSWIK-----DGYCDKACNNSACDWDGDCSGNSGGSKY 477
Db 1367 WTGPRCEAPAAPEVSEPRCPRAACQAKRGDQRCDRECNSPGCGWDGDC-----1417
QY 478 IAGGGGTSGIGVGHWPFGGIN-----SVSYCNQGCAN-----511
Db 1418 -----SLVGDPRQCEALQCWLRFNNSRCDPACSPACLYDNFDCHAGGRERTCNP 1469
QY 512 ----SWLADKF-----CDQACNVLSGCFDAGDGGQD 538
Db 1470 VZEKYCADHFADGRCDQGCNTEECGWDGLDCASE 1503

RESULT 15
T09059
notch4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09059
R:Rowen L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Sc
submitted to the EMBL Data Library, October 1997
A:Description: Sequence of the mouse major histocompatibility locus class III region.
A:Reference number: Z16543
A:Accession: T09059
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1964 <ROW>
A:Cross-references: UNIPROT:P31695; EMBL:AF030001; NID:G2564945; PID:G2564947
C:Genetics:
A:Gene: notch4
A:Map position: 17
A:Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67
1679/3; 1729/1; 1761/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: receptor; signal transduction
F:514-545/Domain: EGF homology <EGF>

Query Match 3.1%; Score 152; DB 2; Length 1964;
Best Local Similarity 28.1%; Pred. No. 0.76;
Matches 41; Conservative 11; Mismatches 32; Indels 62; Gaps 8;
QY 434 WVPVNC-----AEGCPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSKYIAGGGGTSGIGV 489
Db 1162 PGPRCQRPAGSGCEGRG-GDGTCDAGCGSGPGGDWDGDC-----SLGV 1203
QY 490 GHPWQFGGGINSVSYC-----NQCCA-----NSWLADKF- 518
Db 1204 PDFWK---GCPFHQCWLLFRDGRCHPCQDSECLFDGYDCEIPLTCIPAYDQYCRDHFH 1260
QY 519 ---CDQACNVLSGCFDAGDC---QQD 538
Db 1261 NGHCEKGCNNAECGWDGDCRPEGED 1286

Search completed: November 21, 2004, 12:56:00
Job time : 30.3503 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:46:58 ; Search time 133.057 Seconds
(without alignments)
4012.911 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907

Sequence: 1 MLFKLLQRQTYTCLSHRYGL.....SLKTQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4473	91.2	847	Q96N13	Q96N13 homo sapien
2	3314	67.5	950	Q9UL12	Q9UL12 homo sapien
3	2565.5	52.3	921	Q6P9R5	Q6P9R5 mus musculus
4	2565.5	52.3	921	AAH60638	AAH60638 mus muscu
5	2495	50.8	450	Q6IPW5	Q6IPW5 homo sapien
6	2495	50.8	490	AAH71687	AAH71687 homo sapi
7	837	17.1	490	Q86TQ2	Q86TQ2 homo sapien
8	734.5	15.0	384	Q61340	Q61340 mus musculus
9	488	9.9	666	Q9V553	Q9V553 drosophila
10	488	9.9	666	Q8SX14	Q8SX14 drosophila
11	486.5	9.9	717	Q7Q098	Q7Q098 anopheles g
12	360	7.3	71	Q9BUA5	Q9BUA5 homo sapien
13	329	6.7	602	Q9L112	Q9L112 streptomyc
14	324.5	6.6	942	Q9L114	Q9L114 streptomyc
15	322.5	6.6	586	Q69853	Q69853 streptomyc
16	320.5	6.5	486	Q7X4S1	Q7X4S1 neisseria m
17	314	6.4	541	Q69851	Q69851 streptomyc
18	287	5.8	238	Q8GPD3	Q8GPD3 streptococc
19	287	5.8	238	Q9EVX1	Q9EVX1 streptococc
20	266	5.4	545	Q83U59	Q83U59 neisseria m
21	266	5.4	545	Q84D00	Q84D00 neisseria m
22	266	5.4	545	Q9JW88	Q9JW88 neisseria m
23	265	5.4	545	Q68215	Q68215 neisseria m
24	260.5	5.3	442	Q86TW6	Q86TW6 dictyostell
25	259	5.3	545	Q84CZ9	Q84CZ9 neisseria m
26	258	5.3	333	Q8GP72	Q8GP72 streptococ
27	258	5.3	483	Q86KX0	Q86KX0 dictyostell
28	256.5	5.2	364	Q714U9	Q714U9 haemophilus
29	256.5	5.2	364	AAQ12659	AAQ12659 haemophil
30	256	5.2	532	Q06628	Q06628 mycobacteri
31	256	5.2	532	Q7U184	Q7U184 mycobacteri

ALIGNMENTS

RESULT 1

Q96N13

ID Q96N13 PRELIMINARY; PRT; 847 AA.

AC Q96N13;

DT 01-DEC-2001 (TREMELrel. 19, Created)

DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)

DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)

DE Hypothetical protein FLJ151575.

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX NCBI_TaxID=9606;

NP [1]

RP SEQUENCE FROM N.A.

RX PubMed=14702039;

RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,

RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

RA Sekine M., Ohtsubashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,

RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,

RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,

RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,

RA Abe K., Kamiyama K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,

RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,

RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,

RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hottuta T., Kusano J.,

RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,

RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Notsuka S.,

RA Yoshikawa Y., Matsunawa H., Ichihara T., Shichata N., Sano S.,

RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,

RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,

RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,

RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,

RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara Y.,

RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

RA Togashi K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Nakai K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yanashita R.,

RA Okumura K., Nagase T., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,

RT "Complete sequencing and characterization of 21,243 full-length human

RT cDNAs";

RL Nat. Genet. 36:40-45 (2004).

DR EMEL; AK056137; BAB71102.1;

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR010506; DMAP-binding.

DR InterPro; IPR000800; DMAP-binding.

DR Pfam; PF08464; DMAP-binding; 1.

DR SMART; SM00066; Notch; 1.

DR SMART; SM00004; NL; 2.

Q50025 mycobacteri
Q84chl actinobacil
Q8ksb4 actinobacil
Q86hr4 dictyostell
Q88xj7 lactobacilli
Q83vr8 streptococc
Q61q5 streptococc
Q61s6 streptococc
Q84br7 aeromonas h
Q9gr0 neisseria m
Q51151 neisseria m
Q84bk9 aeromonas h
Q7rsw0 giardia lam
Q90285 carassius a

32 255.5 5.2 542 2 Q50025
33 255 5.2 370 2 Q84chl
34 254 364 2 Q8ksb4
35 252 5.1 337 2 Q86hr4
36 249 5.1 337 2 Q88xj7
37 243 5.0 333 2 Q83vr8
38 240 4.9 333 2 Q61q5
39 237 4.8 332 2 Q61s6
40 231.5 4.7 364 2 Q84br7
41 230.5 4.7 366 2 Q9gr0
42 228.5 4.7 373 2 Q51151
43 226 4.6 360 2 Q84bk9
44 206.5 4.2 363 2 Q7rsw0
45 191 3.9 1476 2 Q90285

FT NON TER 847 847
SQ SEQUENCE 847 AA; 95165 MW; 0A11501582473305 CRC64;

Query Match 91.2%; Score 4473; DB 2; Length 847;
Best Local Similarity 99.8%; Pred. No. 1.1e-269;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRVGLVYCFVGVVTVISAFQGEVWLSRQDYHVLFDSDYRNI 60
DB 1 MLFKLLQRTYTCLSHRVGLVYCFVGVVTVISAFQGEVWLSRQDYHVLFDSDYRNI 60

QY 61 AGSKFQNLRLCLPMPIDVYTVWNGTDLLELLKELQVREOMBEEOKAMREILGNKNTTEPTK 120
DB 61 AGSKFQNLRLCLPMPIDVYTVWNGTDLLELLKELQVREOMBEEOKAMREILGNKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLPSPFHSASDIFNVAKPNPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLPSPFHSASDIFNVAKPNPSTNV 180

QY 181 SVVWFDSKVEDAHSGLLKNSQRTVWRGYLTDDKEVPGVLVLMODLAFLSGPPPTFKET 240
DB 181 SVVWFDSKVEDAHSGLLKNSQRTVWRGYLTDDKEVPGVLVLMODLAFLSGPPPTFKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOTKKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOTKKNMTIDGKELTISPA 300

QY 301 YLLWLSAISQSKQDEDSASRPFEDNEELRYSLRSHAPVWRNIFVITNGQIPSWLNLPRT 366
DB 301 YLLWLSAISQSKQDEDSASRPFEDNEELRYSLRSHAPVWRNIFVITNGQIPSWLNLPRT 366

QY 361 DNPRVTIVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVWFGKVPDPDF 420
DB 361 DNPRVTIVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVWFGKVPDPDF 420

QY 421 YSHSGQKQYLTWPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480
DB 421 YSHSGQKQYLTWPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480

QY 481 GGGTSGIGVHPWPGGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTSGIGVHPWPGGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHF 540

QY 541 HELYKVIILLPQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTHL 600
DB 541 HELYKVIILLPQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTHL 600

QY 601 IWHSGMATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660
DB 601 IWHSGMATTIHFNLTFONTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLP 660

QY 661 EAEILFEDIPKPKPPKRDVNSTRAAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPKPKPPKRDVNSTRAAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDITLKGYNLSKALLRSFLMNSOHAQIKNOAIITDETNDLSVAPQEKVHKSILNSLG 780
DB 721 GDITLKGYNLSKALLRSFLMNSOHAQIKNOAIITDETNDLSVAPQEKVHKSILNSLG 780

QY 781 VSELRQRTFFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTEKPPSLIV 840
DB 781 VSELRQRTFFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTEKPPSLIV 840

QY 841 PLESQMT 847
DB 841 PLESQMT 847

RESULT 2
Q9ULL2
ID Q9ULL2
AC Q9ULL2;
DT 01-MAY-2000 (Tremblrel. 13, Created)

01-OCT-2002 (Tremblrel. 22, Last sequence update)
01-MAR-2004 (Tremblrel. 26, Last annotation update)
KIAA1208 protein (Fragment).
Name=KIAA1208;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; Pubmed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345 (1999).
DR EMBL: AB033034; BA866322.2;
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR GO: GO:0030154; P:cell differentiation; IEA.
DR InterPro: IPR010506; DMAP binding.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000800; Notch region.
DR Pfam: PF06464; DMAP binding; 1.
DR Pfam: PF00036; ezh2; 1.
DR Pfam: PF00066; Notch; 1.
DR SMART: SM00004; NL; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
FT NON TER 1 1
SQ SEQUENCE 950 AA; 108785 MW; F653BDFP6AEAC7503 CRC64;

Query Match 67.5%; Score 3314; DB 2; Length 950;
Best Local Similarity 99.8%; Pred. No. 1.9e-197;
Matches 621; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 307 SAISQSKQDEDSASRPFEDNEELRYSLRSHAPVWRNIFVITNGQIPSWLNLPRT 366
DB 1 SAISQSKQDEDSASRPFEDNEELRYSLRSHAPVWRNIFVITNGQIPSWLNLPRT 60

QY 367 IVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVWFGKVPDPDFYSHSG 426
DB 61 IVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVWFGKVPDPDFYSHSG 120

QY 427 QKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGS 486
DB 121 QKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGS 180

QY 487 IGVGHPWPGGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHFELYKV 546
DB 181 IGVGHPWPGGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGDCQDHFELYKV 240

QY 547 ILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTHLHSGM 606
DB 241 ILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTHLHSGM 300

QY 607 NATTIHNLTFONTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLPBAEILF 666
DB 301 NATTIHNLTFONTNDEEFKQITVEVDTRGPKLNSTAGKYENLVSPITLLPBAEILF 360

QY 667 EDIPKPKPPKRDVNSTRAAEVKIPLVNISSLPKDAQLSLNTLDLQLEHGDITLK 726
DB 361 EDIPKPKPPKRDVNSTRAAEVKIPLVNISSLPKDAQLSLNTLDLQLEHGDITLK 420

QY 727 GYNLSKALLRSFLMNSOHAQIKNOAIITDETNDLSVAPQEKVHKSILNSLGVSERLQ 786
DB 421 GYNLSKALLRSFLMNSOHAQIKNOAIITDETNDLSVAPQEKVHKSILNSLGVSERLQ 480

QY 787 RLTFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTEKPPSLIVPLESQM 846
DB 481 RLTFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTEKPPSLIVPLESQM 540

QY 847 TREKKTITGKEKNSRMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYFQDLDLDE 906

```

Db 541 TKEKTKCKENSRMEENAHNGVTEVLLGRKLQHYTSLYGLFWPKKKYFQDLDLDE 600
QY 907 BESLKTQLAYTDSKNTGRQLK 928
Db 601 BESLKTQLAYTDSKNTGRQLK 622

RESULT 3
QEP9R5 PRELIMINARY; PRT; 921 AA.
ID Q6P9R5
AC Q6P9R5;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Woxley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060638; AAH60638.1; -.
DR InterPro; IPR002048; EF-Hand.
DR InterPro; IPR010983; EF-Hand like.
DR Pfam; PF00036; efold; 1.
DR Pfam; PF00066; Notch; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 52.3%; Score 2565.5; DB 2; Length 921;
Best Local Similarity 61.4%; Pred. No. 7.6e-151;
Matches 505; Conservative 35; Mismatches 53; Indels 229; Gaps 6;

QY 107 MREILGNTEPTKSKSQECLLTHCIVKPMVLDPALPANTLKDVPSPSPHSASD 166
Db 1 MREILGNTEPTKSKSQECLLTHCIVKPMVLDPALPANTLKDVPSPSPHSASD 60
QY 167 IFNVAKPNSTNVVVVFDSTKDVEDAHGSLKGNSTQVWVGYLTTDXEVPGLVLMQD 236
Db 61 MFNVAKPNSTNVVVVFDSTKDVEDAHGAPFGKSGKQWVWVAYLTTDXEVPGLVLMQ 120
QY 227 LAFUSGPPPTFKETNQLKTKLPENLSKVKLLQLYSEASVALLKLNPNKDFQBLNKQTK 286

```

```

Db 121 LAFUSGPPPTFKETNQLKTKLPENLSKVKLLQLYSEASVALLKLNPNKDFQBLNKQTK 180
QY 287 NMTIDGKELTISPAYLLWDLSSAISQSKQDEDSASRFEDNEELRYSLRSIERHAPVWRI 346
Db 181 NMTIDGKELTISPAYLLWDLSSAISQSKQDEDSASRFEDNEELRYSLRSIERHAPVWRI 240
QY 347 FIVTNGQIPSWNLNDNPRVTIVTHQDVFENLHSLPTFSSPAIESHIHRIEGLSOKFIYLN 406
Db 241 FIVTNGQIPSWNLNDNPRVTIVTHQDVFENLHSLPTFSSPAIESHIHRIEGLSOKFIYLN 300
QY 407 DDMFGKDWPDPPFVSHSGKQVYLTWPVPCNCAEGCPGKSWIKDGYCDKACNNSACDWDGG 466
Db 301 DDMFGKDWPDPPFVSHSGKQVYLTWPVPCNCAEGCPGKSWIKDGYCDKACNNSACDWDGG 360
QY 467 DCSNGSGSRVYTAGGGTSGIGVHPWPGGGINSVYCNQCCANSWLADKFCDDACNVL 526
Db 361 DCSNGTAGNRFVAGGGTGNIGAGHWFQGGINTISYCNQCCANSWLADKFCDDACNVL 420
QY 527 SCGFADAGCGQDHPHELYKVILLPNQTHVIIIPKGCPLPYFSPAEVAKRGVEGAYSNDPII 586
Db 421 SCGFADAGCGQDHPHELYKVILLPNQTHVIIIPKGCPLPYFSPAEVAKRGVEGAYSNDPII 464
QY 587 REASTANKWKTHILMHSGMNATTIHPNLTFTONTDEEFKMOITVEVDTREGPKLNSTAQ 646
Db 465 -----
QY 647 KGVENLVSPITLLPEAEILFEDIPEKRPFPKRDHVNSTREAQBEVKIPLVNISLLPKD 706
Db 465 -----
QY 707 AQLSLNTLDLQLEHGDIITLKGYNLSKALLRSFLNNSHAKIKNOAIITDETNDLSVAPQ 766
Db 465 -----
QY 767 EKQVHKSILPNSLGVSERLQRLTFPAVSVMVNGHCGQGNPPDLDTTARFVETHTQKTI 826
Db 467 -----
QY 827 GGNVTKEKPPSLIVPLESOMTEKKITGKEKNSMEENAHNGVTEVLLGRKLQHYTD 886
Db 505 GVTVSKENSLPIVPPESHLP-----KEESDRAEGNA---VPVKELVPPRRILQ---Q 551
QY 887 SYLGFPLPWKKYFQDLDDEESLTKQLAYFTDSKNTGRQLK 928
Db 552 NYPGFPLPWKKYFQDLDDEESLTKQLAYFTDSKNTGRQLK 593

RESULT 4
AAH60638 PRELIMINARY; PRT; 921 AA.
ID AAH60638
AC AAH60638;
DT 02-VAR-2004 (TREMELrel. 27, Created)
DT 02-VAR-2004 (TREMELrel. 27, Last sequence update)
DT 02-VAR-2004 (TREMELrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060638; AAH60638.1; -
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D978B CRC64;

Query Match 52.3%; Score 2565.5; DB 2; Length 921;
Best Local Similarity 61.4%; Pred. No. 7.6e-151;
Matches 505; Conservative 35; Mismatches 53; Indels 229; Gaps 6;

QY 107 MRETLGNTTPEPKKSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPFHSASD 166
DB 1 MRETLGNTTPEPKKSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPFHSASD 60

QY 167 IFNAKPNSTNVSVVVFSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMDQ 226
DB 61 MFNAKPNSTNVSVVVFSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMDQ 120

QY 227 LAFSGPPTPKETNOLTKLLENLASKVKLLQYSEASVALLKLNPNDFQBLNKTQK 286
DB 121 LAFSGPPTPKETNOLTKLLENLASKVKLLQYSEASVALLKLNPNDFQBLNKTQK 180

QY 287 NMTIDGKELTSPAYLWDLWSAISQKQEDISASRFENEELRYSLRISIERHAPVRNI 346
DB 181 NMTIDGKELTSPAYLWDLWSAISQKQEDISASRFENEELRYSLRISIERHAPVRNI 240

QY 347 FIVTNGQIPSWLNDPRVTIVTHQDFVRLNLSLPTFSSPAIESHTRIEGSKQFIYN 406
DB 241 FIVTNGQIPSWLNDPRVTIVTHQDFVRLNLSLPTFSSPAIESHTRIEGSKQFIYN 300

QY 407 DDVNFGRDVPDDFYSHSKGQKYLWTPVPCNAEGCGPFWIKDGYCDKACNNSACDWDG 466
DB 301 DDVNFGRDVPDDFYSHSKGQKYLWTPVPCNAEGCGPFWIKDGYCDKACNNSACDWDG 360

QY 467 DCSNCGSRYIAGGGTSGVGHFPHQFGGINSYCNQGCANSLADKCDQACNVL 526
DB 361 DCSNCGSRYIAGGGTSGVGHFPHQFGGINSYCNQGCANSLADKCDQACNVL 420

QY 527 SCGFDAGCGQDHFHLYKVLIPNTHYIIPKGECLIPFSPFAVAKRGVEGAYSNDPII 586
DB 421 SCGFDAGCGQDHFHLYKVLIPNTHYIIPKGECLIPFSPFAVAKRGVEGAYSNDPII 464

QY 587 RHASIANKWKTIHLIMSGNNTTIHNLFTQNTNDEEFKQQTVEVDIREGPKLNSTAQ 646
DB 465 -----BH----- 464

QY 647 KGVENLVSPITLLPEABILFEDIPKRPFKRKHVDNSTRRAQEEVKIPLVNI SLLPKD 706
DB 465 ----- 464

QY 707 AQLSANTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDENSDLSVAQ 766
DB 465 -----BH----- 466

QY 767 EKQVHKSLPNSLGVSERLORLTTPAVSVKNGHDQGNPPLDLETTARFRVETHTKTI 826
DB 467 -----RSEWTAPETVTVKGRDHAINPPPVLETNARL-----AQPTL 504

QY 827 GGNVTKSKPSPSLVPLESQNTKEKKTITGSKENSRMEENAEHIGVTEVLLGRKLOHYTD 886

DB 505 GVTYSKNSLPLVPPSHLP-----KEESDRAGNA---VPVKELVPGRELQ---Q 551
QY 887 SYLGLFWKKKKYFQDLDDEESLKTQLAYFTDSKNTGROLK 928
DB 552 NYPGLFWKKKKYFQDLDDEESLKTQLAYFTDSKNTGROLK 593

RESULT 5
QIPWS PRELIMINARY; PRT; 490 AA.
AC Q6IPW5;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE MG4170 protein.
GN Name=MG4170;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071687; AAH71687.1; -
DR InterPro; IPR000800; Notch_region.
DR Pfam; PF00066; Notch; 1.
DR SMART; SM00004; NL; 1.
SQ SEQUENCE 490 AA; 55982 MW; 9223139711337ED0 CRC64;

Query Match 50.8%; Score 2495; DB 2; Length 490;
Best Local Similarity 99.8%; Pred. No. 7.2e-147;
Matches 469; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFLLQRTYTCLSHRYGLVCFVGVVTVISAFQGEVVLWSRQDQYHVFDSYRNI 60
DB 1 MLFLLQRTYTCLSHRYGLVCFVGVVTVISAFQGEVVLWSRQDQYHVFDSYRNI 60

QY 61 AKGSPQNRCLPMPIDVYVTVWNGTDLELLKLOOVRQMEEEKANREILKNTTPTK 120
DB 61 AKGSPQNRCLPMPIDVYVTVWNGTDLELLKLOOVRQMEEEKANREILKNTTPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLYPSFHSASDIFNAKPNPSTNV 180
DB 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLYPSFHSASDIFNAKPNPSTNV 180

QY 181 SVVVFSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMDQLAFLSGFPPTKET 240

```

Db 181 SVVFDSTKDVDAHSGLLKGNRSQTVMRGVLTDDKEVGLVMDLAFSLGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
QY 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
QY 361 DNPRTVITHQDVFRNLSHLPTFSSPAIESHIHRIEGLSKQFIYLNDDVMFGKDVWPDFF 420
Db 361 DNPRTVITHQDVFRNLSHLPTFSSPAIESHIHRIEGLSKQFIYLNDDVMFGKDVWPDFF 420
QY 421 YHSKGGQKVLTPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSG 470
Db 421 YHSKGGQKVLTPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSG 470

```

RESULT 6

```

AAH71687 ID AAH71687 PRELIMINARY; PRT; 490 AA.
AC AAH71687;
DT 01-JUN-2004 (TremBLrel. 27, Created)
DT 01-JUN-2004 (TremBLrel. 27, Last sequence update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]

```

SEQUENCE FROM N.A.

```

TSSUP=Brain;
MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J.,
RA Bosak S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madañ A., Rodríguez S., Sanchez A.,
RA Whiting M., Madañ A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]

```

SEQUENCE FROM N.A.

```

TSSUP=Brain;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071687; AAH71687.1; -
KW Hypothetical protein.
SQ SEQUENCE 490 AA; 55982 MW; 9223139711337ED0 CRC64;

```

```

Query Match 50.8%; Score 2495; DB 2; Length 490;
Best Local Similarity 99.8%; Pred. No. 7.2e-147;
Matches 469; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVTIISAFQFGEVLEWSDQHVLVFDSDYRNI 60
Db 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVTIISAFQFGEVLEWSDQHVLVFDSDYRNI 60

```

```

QY 61 AGSKFQNRICLPMPIDVWYTWNGTDLLELLKELQOVREQMBEEOQAMREILGNTEPTK 120
Db 61 AGSKFQNRICLPMPIDVWYTWNGTDLLELLKELQOVREQMBEEOQAMREILGNTEPTK 120
QY 121 KSEKQLECLLTHCICIKVPMVLVDPALPANITLKDPSLYSPFSHASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCICIKVPMVLVDPALPANITLKDPSLYSPFSHASDIFNVAKPKNPSTNV 180
QY 181 SVVFDSTKDVDAHSGLLKGNRSQTVMRGVLTDDKEVGLVMDLAFSLGFPPTFKET 240
Db 181 SVVFDSTKDVDAHSGLLKGNRSQTVMRGVLTDDKEVGLVMDLAFSLGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300
QY 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
Db 301 YLLWDLISAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360
QY 361 DNPRTVITHQDVFRNLSHLPTFSSPAIESHIHRIEGLSKQFIYLNDDVMFGKDVWPDFF 420
Db 361 DNPRTVITHQDVFRNLSHLPTFSSPAIESHIHRIEGLSKQFIYLNDDVMFGKDVWPDFF 420
QY 421 YHSKGGQKVLTPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSG 470
Db 421 YHSKGGQKVLTPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSG 470

```

RESULT 7

```

088TQ2 ID 088TQ2 PRELIMINARY; PRT; 490 AA.
AC 088TQ2;
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Similar to v-maf musculoaponeurotic fibrosarcoma oncogene family,
DE protein B (Avian) (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]

```

SEQUENCE FROM N.A.

```

RP TISSUE=Liver;
RC Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042615; AAH42615.1; -
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR010506; DMAP binding.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF Hand like.
DR Pfam; PF06464; DMAP_binding; 1.
DR Pfam; PF00036; ethand; 1.
DR PROSITE; PS00018; EF HAND; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 490 AA; 57411 MW; 37B19FE0D1259AD2 CRC64;

```

```

Query Match 17.1%; Score 837; DB 2; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.3e-43;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 767 EKQVHKSILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLLETTARFVETHTKTI 826
Db 1 EKQVHKSILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLLETTARFVETHTKTI 60

```

```

QY 827 GGNVTKEKPPSLIPLSQMTKEKKTGKENSRMENAEHNHGVTEVLLGRKLQHYTD 886
Db 61 GGNVTKEKPPSLIPLSQMTKEKKTGKENSRMENAEHNHGVTEVLLGRKLQHYTD 120

```

```

QY 887 SYLGFLPWEKKKYFQDLDDEESLKTQLAYPTDSKNTGRQLK 928
Db 121 SYLGFLPWEKKKYFQDLDDEESLKTQLAYPTDSKNTGRQLK 162

```



```

RESULT 8
Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Basic domain/leucine zipper transcription factor (fragment).
GN Name=Maif;
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95094266; PubMed=8001130;
RX Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RL zipper transcription factor.";
RL Cell 79:1023-1034(1994).
DR EMBL; L36434; AAA5688.1; -.
DR PIR; I49528; I49528.
DR MG; MG1:104555; Maif.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0043544; P:positive regulation of transcription from P. .; IDA.
DR GO; GO:0043544; P:positive regulation of transcription from P. .; IDA.
DR GO; GO:0007379; P:segment specification; IMP.
DR InterPro; IPR010506; DMAP binding.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF06464; DMAP binding; 1.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT NON TER
SQ SEQUENCE 384 AA; 4375 MW; 51F473C6807A7B55 CRC64;

Query Match 15.0%; Score 734.5; DB 2; Length 384;
Best Local Similarity 57.9%; Pred. No. 2.2e-37;
Matches 165; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

Qy 644 TAQGYENLVSPITLLPEABILEDPKPKFKPKHVNSTRPAQEVKIPLVNISLL 703
Db 3 TTQAYESLVSPVTPLPQADVPPEDVPKPKFKPKHVNSTRPAQEVKIPLVNISLL 62

Qy 704 PKDAQLSINTLDLQLEHGDITLKGYNLSKALLRSFLMNSOHAKIKNOALITDETNDLSLV 763
Db 63 PKAQVRLSNDLQLEKGDITLKGYNLSKALLRSFLGNSLDTKIKPQA-RIDETKGNLE 121

Qy 764 APQKQVHKHSILPNSLGVSRLOLTPPAVSVKNGHDOQNPPDLDTTARFRVETHQ 823
Db 122 VPQENPSHRR--PHGFAGEHRSERWTAPATVTVKGRDHALNPPVLETNARL-----AQ 174

Qy 824 KTIQNTVKPPSLIIVPLESQMTKEKIKTKGKENSRENEAENHIGTVTLVLRKLQH 883
Db 175 PTLGVTVSKENLSPLIIVPPESHLP-----KSESDRAEGNA---VPVKELVPGRCRSR 224

Qy 884 YTSYLGFLPEWKKKYFQDLDEESLKTOLAYFTDSKNTGRQLK 928
Db 225 IIQA---FCGKKKKYQDLDAEESLKTOLAYFTDSKNTGRQLK 266

RESULT 9
Q9V553 PRELIMINARY; PRT; 666 AA.
AC Q9V553;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

```

```

DE GN ORFNames=CG8027;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.B.D.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Poillard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J., Ye J.,
RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2183-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426065; PubMed=12537568;
RX Celnik S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RL melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnik S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RL a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426069; PubMed=12537572;
RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

```



```
Db 490 NPRERLAVDAEQTR-VKRLAGAYEG---KAYVAELLKERTHAPGVLLAERLEAVEGEVA 545
QY 229 FLSGPPTFKETNQLK-----TKLPENLSK-----VKLLQLYSE 263
Db 546 GLAVREPVVTSRTRLRGPAYGCDIEFRQVPEEGDGFVAPLRPSAVGPKPLSLTPD 605
QY 264 ASVALLKLNPKDFOELNKOTKK-----NMTIDGKELTISPAYLLWD-----LSAI 309
Db 606 ARTEV-----KOREYPTLEPTKLVSDITFPDAVTVWDDSDPRNQRARRAALGLE 661
QY 310 SQSKQBEDISAFEDNEELRYSLRSTIERHAPVWRNIFIVTNGQIPSWLNLNPRVTIYT 369
Db 662 AESSGDE--AARFRNRDELRYSLRSLAMPAPWIRKIYLVTDQTPWLNTEHEGIEVVS 718
QY 370 HQDVERNLSHLPTFSSPAIESHRIEGLSQKLYLNDVDFGKDYWPDFFYSHSKGQV 429
Db 719 HRDIFTDQDCLPFTFNSHSIESQLHHDIGSEQFLYLNDDVFICRPVGAORFFLPNGASRF 778
QY 430 YLTWPVPCNAEGCGPGWIKDGYCDKACNNSA 460
Db 779 F--WSPTTVPVGEP-TEEDSGYFAAAKNNA 806
```

RESULT 15

```
O69853 PRELIMINARY; PET; 586 AA.
AC O69853;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein SCO6023.
GN OrderedLocusNames=SCO6023; ORFNames=SCIC3.11;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939126; CAA19235.1; -.
DR PIR; T34703; T34703.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 586 AA; 66009 MW; 75F98F0916484116 CRC64;
```

```
Query Match
Best Local Similarity 46.3%; Pred. No. 1.9e-11;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;
QY 290 IDGKELTISPAYLLW--DLSAISQSKQDEDISAFEDNEELRYSLRSTIERHAPVWRNIF 347
Db 282 VQND-----PA---WRRRSAYGGYHANSANARYISDELRYSLRALEQNAQAPVRRVHV 334
QY 348 IVTNGQIPSWLNLNPRVTIHTQDVFRNLSHLPTFSSPAIESHRIEGLSQKLYLND 407
Db 335 LVTDGQRPAWLNDSHRPLTVVDHSEIFADPAALPTFNSHAIESRLHHIKGLSEHFLYLN 394
QY 408 DVWFGKDVWPDFFY 421
Db 395 DMFLGRPVTPQDFF 408
```

Search completed: November 21, 2004, 13:02:00
Job time : 135.057 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:38:18 ; Search time 37,6766 Seconds
(without alignments)
3122.980 Million cell updates/sec

Title: US-10-023-888-5

Perfect score: 1731

Sequence: 1 DTPADSLRYVVKILNSKFGF.....RKIFPRRIKESPNRIIV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	ABR61377 Human Glc
2	1731	100.0	328	7	ADD27813 Human Glc
3	1731	100.0	328	7	ABW01489 Human Glc
4	1731	100.0	328	7	ABW01538 Human Glc
5	1731	100.0	1196	5	AAE25294 Human nuc
6	1731	100.0	1256	5	AAE25290 Human nuc
7	1698	98.1	328	4	ABR61380 Murine Glc
8	1698	98.1	328	7	ADD27818 GlcNAc-ph
9	1698	98.1	328	7	ABW01492 Mouse pro
10	1698	98.1	328	7	ABW01541 Mouse pro
11	1487	85.9	1199	7	ADD27810 Soluble h
12	1487	85.9	1199	7	ABW01487 N-acetylgl
13	1487	85.9	1199	7	ABW01536 N-acetylgl
14	1276	73.7	248	7	ADG39858 Protein s
15	1140	65.9	663	7	ADG39855 Protein s
16	1140	65.9	1459	6	ABU07381 Human pro
17	1140	65.9	1459	7	ADG39794 Human pro
18	704	40.7	132	4	AAE93369 Human pro
19	581	33.6	384	7	ADG39857 Protein s
20	506	29.2	652	4	ABE59094 Drosophil
21	495.5	28.6	651	7	ADG39859 Protein s
22	402	23.2	502	4	ABR61383 Partial D
23	402	23.2	502	7	ADD27824 Fruit fly
24	402	23.2	502	7	ABW01495 Partial f
25	402	23.2	502	7	ABW01544 Partial f

26	300	17.3	55	8	ABO57141 Human gen
27	145	8.4	28	4	ABR61390 Bovine Gl
28	145	8.4	28	4	AAE68261 Amino ter
29	124	7.2	77	4	AAE88395 Human imm
30	124	7.2	1312	5	ABR77984 Amino aci
31	124	7.2	1312	6	ABR53887 Protein s
32	124	7.2	1312	7	ADK63602 Disease t
33	113.5	6.6	545	2	AAW79296 Neisseria
34	112	6.5	46	7	ADD27834 GlcNAc-ph
35	106	6.1	574	5	ABP26069 Streptoco
36	101.5	5.9	301	5	ABG49330 Listeria
37	101	5.8	2633	4	ABQ06505 Novel hum
38	101	5.8	2663	8	ADQ17932 Human sof
39	101	5.8	2688	4	AAE40883 Human poi
40	100.5	5.8	259	2	AAE79314 IL-3 cont
41	100.5	5.8	259	2	AAE79331 IL-3 cont
42	100.5	5.8	259	3	AAE53212 Human int
43	100.5	5.8	259	3	AAE53195 Human int
44	100.5	5.8	259	4	AAE14006 Myeloipoie
45	100.5	5.8	259	4	AAE13988 Myeloipoie

ALIGNMENTS

RESULT 1

ABR61377

ID ABR61377 standard; protein; 328 AA.

XX AC ABR61377;

XX DT 01-AUG-2003 (first entry)

XX DE Human GlcNAc-phosphotransferase beta-subunit.

XX KW Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;

XX KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;

XX KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;

XX KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;

XX KW lysosomal storage disease; enzyme; beta-subunit.

XX OS Homo sapiens.

XX PN US6537785-B1.

XX PD 25-MAR-2003.

XX PF 10-AUG-2000; 2000US-00636077.

XX PR 14-SEP-1999; 99US-0153831P.

XX PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX PI Canfield WM;

XX PT WPI; 2001-290356/30.

XX DR N-PSDB; ACCB1001.

XX PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX PS Disclosure; Page 28-29; 62pp; English.

XX CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has

XX CC Acetylglucosaminidase) (II). The protein of the invention has

XX CC nephrotropic activity, and may be useful in enzyme replacement therapy. A

XX CC protein of the invention (I), (II) is useful for preparing a

XX CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase

XX CC comprising a terminal mannose-6-phosphate, is useful for treating a

CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention
 XX
 SQ Sequence 328 AA;
 Query Match 100.0%; Score 1731; DB 4; Length 328;
 Best Local Similarity 100.0%; Pred. No. 7.3e-160; Indels 0; Gaps 0;
 Matches 328; Conservative 0; Mismatches 0;
 QY 1 DTFADSLRVVVKILNSKFGFTSRKVPAPMPHMDIRVMOELQDMFPPEFDKTSFHKVRHS 60
 DB 1 DTFADSLRVVVKILNSKFGFTSRKVPAPMPHMDIRVMOELQDMFPPEFDKTSFHKVRHS 60
 QY 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNPVTKSLVNTCKEPTVDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNPVTKSLVNTCKEPTVDKIHKAYKDK 180
 QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNDKDAQTVKAVL 240
 DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNDKDAQTVKAVL 240
 QY 241 RPFYESMFPISQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300
 DB 241 RPFYESMFPISQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300
 QY 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328

RESULT 2
 ADD27813
 ID ADD27813 standard; protein; 328 AA.
 AC ADD27813;
 XX
 DT 15-JAN-2004 (first entry)
 XX Human GlcNAc-phosphotransferase beta subunit.
 DE
 XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
 KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.
 XX
 OS Homo sapiens.
 XX US2003119088-A1.
 XX
 XX 26-JUN-2003.
 XX
 XX 21-DEC-2001; 2001US-00023888.
 XX
 XX 21-DEC-2001; 2001US-00023888.
 XX
 XX (NOVA-) NOVAZYME PHARM INC.
 XX
 XX Canfield W, Kudo M;
 XX
 XX WPI; 2003-801323/75.
 DR N-PSDB; ADD27813.
 XX
 XX Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.
 XX
 XX Claim 8; SEQ ID NO 5; 55pp; English.
 PS
 XX The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of the human GlcNAc-phosphotransferase beta subunit.
 XX
 SQ Sequence 328 AA;
 Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 7.3e-160; Indels 0; Gaps 0;
 Matches 328; Conservative 0; Mismatches 0;
 QY 1 DTFADSLRVVVKILNSKFGFTSRKVPAPMPHMDIRVMOELQDMFPPEFDKTSFHKVRHS 60
 DB 1 DTFADSLRVVVKILNSKFGFTSRKVPAPMPHMDIRVMOELQDMFPPEFDKTSFHKVRHS 60
 QY 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNPVTKSLVNTCKEPTVDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNPVTKSLVNTCKEPTVDKIHKAYKDK 180
 QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNDKDAQTVKAVL 240
 DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNDKDAQTVKAVL 240
 QY 241 RPFYESMFPISQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300
 DB 241 RPFYESMFPISQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300
 QY 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328

RESULT 3
 ABW01489
 ID ABW01489 standard; protein; 328 AA.
 AC ABW01489;
 XX
 DT 15-JAN-2004 (first entry)
 XX Human GlcNAc-phosphotransferase beta subunit precursor protein.
 DE
 XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
 KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
 KW gastrointestinal; human; enzyme; lectin resistant cell;
 KW deoxymannojirimycin; kifunensine; glycosylation inhibition.
 XX
 OS Homo sapiens.
 XX US2003124652-A1.
 XX
 XX 03-JUL-2003.
 XX
 XX 21-DEC-2001; 2001US-00023889.
 XX
 XX 21-DEC-2001; 2001US-00023889.
 XX
 XX (NOVA-) NOVAZYME PHARM INC.
 XX
 XX Canfield WM;
 XX
 XX WPI; 2003-810984/76.
 DR N-PSDB; RAD62491.
 XX
 XX Producing a high mannose glycoprotein for treating lysosomal storage
 PT disease, comprises culturing the lectin resistant mammalian cell in the
 PT presence of deoxymannojirimycin and kifunensine.
 XX
 XX Claim 10; Page 18-19; 46pp; English.
 PS

XX The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is human N-
 CC acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor
 CC protein
 XX
 SQ Sequence 328 AA;
 Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 7.3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPHPHMDIRIVMQELODMFPEFDKTSFHKVRS 60
 Db 1 DTFADSLRYVNKILNSKFGFTSRKVPAPHPHMDIRIVMQELODMFPEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPLNPPVTKSLVNTCKPVDKIHAYKDK 180
 Db 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPLNPPVTKSLVNTCKPVDKIHAYKDK 180
 QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHDKDAQTVKAVL 240
 Db 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHDKDAQTVKAVL 240
 QY 241 RDFYESMFPPIPSQFELPREYRNRFLLHMLQEWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
 Db 241 RDFYESMFPPIPSQFELPREYRNRFLLHMLQEWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
 QY 301 AEQLIALKRKIFPPRRIRHKEASPNRIRV 328
 Db 301 AEQLIALKRKIFPPRRIRHKEASPNRIRV 328
 RESULT 4
 ABW01538
 ID ABW01538 standard; protein; 328 AA.
 AC ABW01538;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GlcNAc-phosphotransferase beta subunit precursor protein.
 XX
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN US2003124653-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 21-DEC-2001; 2001US-00023890.
 XX
 PR 21-DEC-2001; 2001US-00023890.
 XX
 PA (NOVA-) NOVAZYME PHARM INC.
 XX
 PI Canfield WM;
 XX
 WPI; 2003-810985/76.

DR N-PSDB; AAD62650.
 XX
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 PS Claim 10; Page 18-19; 46pp; English.
 XX
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase beta subunit precursor protein
 XX
 SQ Sequence 328 AA;
 Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 7.3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPHPHMDIRIVMQELODMFPEFDKTSFHKVRS 60
 Db 1 DTFADSLRYVNKILNSKFGFTSRKVPAPHPHMDIRIVMQELODMFPEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPLNPPVTKSLVNTCKPVDKIHAYKDK 180
 Db 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPLNPPVTKSLVNTCKPVDKIHAYKDK 180
 QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHDKDAQTVKAVL 240
 Db 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHDKDAQTVKAVL 240
 QY 241 RDFYESMFPPIPSQFELPREYRNRFLLHMLQEWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
 Db 241 RDFYESMFPPIPSQFELPREYRNRFLLHMLQEWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
 QY 301 AEQLIALKRKIFPPRRIRHKEASPNRIRV 328
 Db 301 AEQLIALKRKIFPPRRIRHKEASPNRIRV 328
 RESULT 5
 AAE25294
 ID AAE25294 standard; protein; 1196 AA.
 XX
 AC AAE25294;
 XX
 DT 30-OCT-2002 (first entry)
 XX
 DE Human nucleic acid-associated protein (NAAP-13).
 XX
 KW Human; nucleic acid-associated protein; NAAP-13; neurological disorder;
 KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
 KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
 KW gene therapy; neurotropic; neuroprotective; cerebroprotective; virucide;
 KW immunosuppressive; protozoaside; antimicrobial.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Peptide Location/Qualifiers
 FT 1..34 /label= Signal_peptide
 FT 1..19 /note= "Cytosolic domain"
 FT 20..42 /note= "Transmembrane domain"
 FT

FT Protein 35. .1196 /note= "Mature human NAAP-13"
FT Domain 43. .1152 /note= "Non-cytosolic domain"
FT Domain 1153. .1175 /note= "Transmembrane domain"
FT Domain 1176. .1196 /note= "Cytosolic domain"
XX WO200250279-A2.
XX 27-JUN-2002.
XX 19-DEC-2001; 2001WO-US050256.
XX 21-DEC-2000; 2000US-0257714P.
XX 05-JAN-2001; 2001US-0260081P.
XX 16-JAN-2001; 2001US-0262302P.
XX 23-JAN-2001; 2001US-0263823P.
XX 02-FEB-2001; 2001US-0266088P.
XX 29-OCT-2001; 2001US-0348442P.
XX (INCY-) INCYTE GENOMICS INC.
XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
XX Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
XX Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
XX Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
XX Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;
XX Swarnakar A, Reddy R;
XX WPI; 2002-519887/55.
XX N-PSDB; AAD41203.
XX Nucleic acid associated proteins and nucleic acids for diagnosing,
XX treating and preventing cell proliferative (e.g. cancers), neurological
XX (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX Claim 68; Page 169-172; 193pp; English.
XX The invention relates to nucleic acid-associated proteins (NAAP) and
XX nucleic acids. The nucleic acid and amino acid sequences are useful for
XX diagnosing, treating and preventing cell proliferative e.g.
XX arteriosclerosis, atherosclerosis, lymphoma or cancers, neurological
XX (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
XX autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
XX (e.g. malaria, or leishmania), as well as in assessing the effects of
XX exogenous compound on the expression of nucleic acid and amino acid
XX sequences of nucleic acid-associated proteins. The invention is useful in
XX gene therapy. The present sequence is human NAAP-13
XX Sequence 1196 AA;
XX Query Match 100.0%; Score 1731; DB 5; Length 1196;
XX Best Local Similarity 100.0%; Pred. No. 4.8e-159;
XX Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMPHMDIRIVNQELQDMFPPEPKTSFHKVRHS 60
Db 869 DTFADSLRYNKNILNSKFGFTSRKVPAPMPHMDIRIVNQELQDMFPPEPKTSFHKVRHS 928
QY 61 EDMQFASFYFLMSAVQPLNISQVDEVDTDQSGVLSDRIRTLARIRHELPLSLQDLT 120
Db 929 EDMQFASFYFLMSAVQPLNISQVDEVDTDQSGVLSDRIRTLARIRHELPLSLQDLT 988
QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIHAYKDK 180
Db 989 GLEHMLNCSKMLPADITQLNNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIHAYKDK 1048
QY 191 NKYRFEIMGEETIAFKMIRTNVSHVGGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240
Db 1049 NKYRFEIMGEETIAFKMIRTNVSHVGGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 1108

QY 241 RDFYESMFPPIPSQFELPREYRNRFLEHMHQEWRAVDKLFKTHCVLATLIMFTFSFF 300
Db 1109 RDFYESMFPPIPSQFELPREYRNRFLEHMHQEWRAVDKLFKTHCVLATLIMFTFSFF 1168
QY 301 AEQIALKRRKIFPRRIHKEASPNRIIV 328
Db 1169 AEQIALKRRKIFPRRIHKEASPNRIIV 1196
RESULT 6
AAE25290
ID AAE25290 standard; protein; 1256 AA.
XX
XX AAE25290;
XX 30-OCT-2002 (first entry)
XX Human nucleic acid-associated protein (NAAP-9).
XX Human; nucleic acid-associated protein; NAAP-9; neurological disorder;
XX arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
XX lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
XX autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
XX gene therapy; nontropic; neuroprotective; cerebroprotective; virucide;
XX immunosuppressive; protozoacide; antimicrobial.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Domain 448. .469 /note= "Notch domain"
XX Domain 500. .536 /note= "Notch domain"
XX Domain 1018. .1030 /note= "EF-hand calcium-binding domain"
XX WO200250279-A2.
XX 27-JUN-2002.
XX 19-DEC-2001; 2001WO-US050256.
XX 21-DEC-2000; 2000US-0257714P.
XX 05-JAN-2001; 2001US-0260081P.
XX 16-JAN-2001; 2001US-0262302P.
XX 23-JAN-2001; 2001US-0263823P.
XX 02-FEB-2001; 2001US-0266088P.
XX 29-OCT-2001; 2001US-0348442P.
XX (INCY-) INCYTE GENOMICS INC.
XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
XX Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
XX Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
XX Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
XX Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;
XX Swarnakar A, Reddy R;
XX WPI; 2002-519887/55.
XX N-PSDB; AAD41199.
XX Nucleic acid associated proteins and nucleic acids for diagnosing,
XX treating and preventing cell proliferative (e.g. cancers), neurological
XX (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX Claim 64; Page 163-165; 193pp; English.
XX The invention relates to nucleic acid-associated proteins (NAAP) and
XX nucleic acids. The nucleic acid and amino acid sequences are useful for
XX diagnosing, treating and preventing cell proliferative e.g.
XX arteriosclerosis, atherosclerosis, lymphoma or cancers, neurological
XX (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
XX autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections

CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-9
XX
SQ Sequence 1256 AA;

Query Match 100.0%; Score 1731; DB 5; Length 1256;
Best Local Similarity 100.0%; Pred. No. 5.1e-159;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMQELODMFPEEPDKTSFHKVRHS 60
DB 929 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMQELODMFPEEPDKTSFHKVRHS 988
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
DB 989 EDMQAFSYFYILMSAVQPLNISQVFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 1048
QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
DB 1049 GLEHMLNCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 1108
QY 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 240
DB 1109 NKYRFEIMGEIEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 1168
QY 241 RDPFESMFPPIPSQFELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
DB 1169 RDPFESMFPPIPSQFELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 1228
QY 301 AEQIALTKRKKIPRRRIHKEASPNRIRV 328
DB 1229 AEQIALTKRKKIPRRRIHKEASPNRIRV 1256

RESULT 7
ABR61380
ID ABR61380 standard; protein; 328 AA.
XX ABR61380;
XX 01-AUG-2003 (first entry)
XX Murine GlcNAc-phosphotransferase beta-subunit.
XX Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW lysosomal storage disease; enzyme; beta-subunit.
OS Mus musculus.
XX US6537785-B1.
XX 25-MAR-2003.
XX 10-AUG-2000; 2000US-00636077.
XX 14-SEP-1999; 99US-0153831P.
XX (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.
XX Canfield WM;
XX WPI; 2001-290356/30.
XX N-PSDB; ACC81007.
XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-
PT phosphodiester alpha-N-Acetylglucosaminidase, useful for producing
PT phosphorylated lysosomal hydrolase for treating lysosomal storage
PT diseases.

XX Disclosure; Page 35-36; 62pp; English.
XX
XX The invention relates to a novel isolated human N-acetylglucosamine-1-
CC phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester
CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-
CC acetylglucosaminidase) (II). The protein of the invention has
CC nephrotropic activity, and may be useful in enzyme replacement therapy. A
CC protein of the invention (I), (II) is useful for preparing a
CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
CC comprising a terminal mannose-6-phosphate, is useful for treating a
CC patient suffering from a lysosomal storage disease. The present sequence
XX is used in the exemplification of the invention
SQ Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 1.2e-156;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMQELODMFPEEPDKTSFHKVRHS 60
DB 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMQELODMFPEEPDKTSFHKVRHS 60
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLNCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 240
DB 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 240
QY 241 RDPFESMFPPIPSQFELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
DB 241 RDPFESMFPPIPSQFELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
QY 301 AEQIALTKRKKIPRRRIHKEASPNRIRV 328
DB 301 AEQIALTKRKKIPRRRIHKEASPNRIRV 328

RESULT 8
ADD27818
ID ADD27818 standard; protein; 328 AA.
XX ADD27818;
XX 15-JAN-2004 (first entry)
XX GlcNAc-phosphotransferase associated protein #2.
XX mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.
XX Mus musculus.
XX US2003119088-A1.
XX 26-JUN-2003.
XX 21-DEC-2001; 2001US-00023888.
XX 21-DEC-2001; 2001US-00023888.
XX (NOVA-) NOVAZYME PHARM INC.
XX Canfield W, Kudo M;
XX WPI; 2003-801323/75.

XX Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.
 XX
 XX Disclosure; SEQ ID NO 10; 55pp; English.
 XX
 XX The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (GDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of a GlcNAc-phosphotransferase associated protein.
 XX
 XX Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 7; Length 328;
 Best Local Similarity 97.6%; Pred. No. 1.2e-156;
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRS 60
 DB 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRS 60
 QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVDKIHKAYKDK 180
 QY 181 NKYRFEIMGEEIEAFKMIKTNTVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDARTVKAVL 240
 DB 181 NKYRFEIMGEEIEAFKMIKTNTVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDARTVKAVL 240
 QY 241 RDFYESMFPPIPSQPELPREYRNRLFMHLEQWRAVDKLFKFWTHCVLATLIMFTTIESFF 300
 DB 241 RDFYESMFPPIPSQPELPREYRNRLFMHLEQWRAVDKLFKFWTHCVLATLIMFTTIESFF 300
 QY 301 AEQIALKRRKIFPRRRHKASPNRIRV 328
 DB 301 AEQIALKRRKIFPRRRHKASPNRIRV 328

RESULT 9
 ABW01492
 ID ABW01492 standard; protein; 328 AA.
 XX
 XX ABW01492;

15-JAN-2004 (first entry)

Mouse protein #2 used to illustrate the method of the invention.

Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
 lysosomal storage disease; gastrointestinal; mouse;
 lectin resistant cell; deoxymannojirimycin; kifunensine;
 glycosylation inhibition.

Mus musculus.

US2003124652-A1.

03-JUL-2003.

21-DEC-2001; 2001US-00023889.

21-DEC-2001; 2001US-00023889.

(NOVA-) NOVAZIME PHARM INC.

PI Canfield WM;
 XX WPI; 2003-810984/76.
 XX
 XX Producing a high mannose glycoprotein for treating lysosomal storage
 PT disease, comprises culturing the lectin resistant mammalian cell in the
 PT presence of deoxymannojirimycin and kifunensine.
 XX
 XX Disclosure; Page 25-26; 46pp; English.

CC The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is mouse protein used to
 CC illustrate the method of the invention

XX Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 7; Length 328;
 Best Local Similarity 97.6%; Pred. No. 1.2e-156;
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRS 60
 DB 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRS 60
 QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVDKIHKAYKDK 180
 QY 181 NKYRFEIMGEEIEAFKMIKTNTVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDARTVKAVL 240
 DB 181 NKYRFEIMGEEIEAFKMIKTNTVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDARTVKAVL 240
 QY 241 RDFYESMFPPIPSQPELPREYRNRLFMHLEQWRAVDKLFKFWTHCVLATLIMFTTIESFF 300
 DB 241 RDFYESMFPPIPSQPELPREYRNRLFMHLEQWRAVDKLFKFWTHCVLATLIMFTTIESFF 300
 QY 301 AEQIALKRRKIFPRRRHKASPNRIRV 328
 DB 301 AEQIALKRRKIFPRRRHKASPNRIRV 328

RESULT 10
 ABW01541
 ID ABW01541 standard; protein; 328 AA.
 XX
 XX ABW01541;

15-JAN-2004 (first entry)

Mouse protein #2 used to illustrate the method of the invention.

Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.

Mus musculus.

US2003124653-A1.

03-JUL-2003.

21-DEC-2001; 2001US-00023890.

```

XX PR 21-DEC-2001; 2001US-00023890.
XX PA (NOVA-) NOVAZYME PHARM INC.
XX PI Canfield W;
XX XX WPI; 2003-810985/76.
XX PT Producing a glycoprotein with reduced complex carbohydrates by culturing
XX PT the lectin resistant mammalian cell expressing the glycoprotein for
XX PT treating lysosomal storage disease.
XX PS Disclosure; Page 25-26; 46pp; English.
XX CC The present invention provides a method of producing a glycoprotein
XX CC having reduced complex carbohydrates by culturing the lectin resistant
XX CC mammalian cell expressing the glycoprotein. The method is useful for
XX CC producing a glycoprotein with reduced complex carbohydrates for treating
XX CC lysosomal storage disease. The present invention is also useful in gene
XX CC therapy. The present sequence is mouse protein used to illustrate the
XX CC method of the invention
XX SQ Sequence 328 AA;
    Query Match 98.1%; Score 1698; DB 7; Length 328;
    Best Local Similarity 97.6%; Pred. No. 1.2e-156;
    Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMPEEFDKTSFHKVRRHS 60
Db 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMPEEFDKTSFHKVRRHS 60
QY 61 EDMQFAFSYFYLYMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSYFYLYMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
QY 121 GLEHMLNCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180
Db 121 GLEHMLNCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180
QY 181 NKYFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDAQTVKAVL 240
Db 181 NKYFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDAQTVKAVL 240
QY 241 RDFYESMFPPIPSQFELPREYRNRFHMHLEQEWRAVDKDKLKFTHCVLATLIIFTIFSFF 300
Db 241 RDFYESMFPPIPSQFELPREYRNRFHMHLEQEWRAVDKDKLKFTHCVLATLIIFTIFSFF 300
QY 301 AEQIIALKRKIFPRRRHKASPNRIRV 328
Db 301 AEQIIALKRKIFPRRRHKASPNRIRV 328

RESULT 11
ADD27810
ID ADD27810 standard; protein; 1199 AA.
XX AC ADD27810;
XX XX 15-JAN-2004 (first entry)
XX DT
XX DE Soluble human GlcNAc-phosphotransferase.
XX KW human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
XX KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
XX KW enzyme.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN US2003119088-A1.
XX XX

```

```

PD 26-JUN-2003.
XX 21-DEC-2001; 2001US-00023888.
XX PR 21-DEC-2001; 2001US-00023888.
XX XX (NOVA-) NOVAZYME PHARM INC.
XX PA Canfield W; Kudo M;
XX PI WPI; 2003-801323/75.
XX DR N-PSDB; ADD27809.
XX XX
XX PT Phosphorylating a protein for treating a patient suffering from a
XX PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
XX PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
XX PT protein.
XX PS Claim 3; SEQ ID NO 2; 55pp; English.
XX CC The invention relates to a method of phosphorylating a protein comprising
XX CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
XX CC acetylglucosamine) and producing a phosphorylated protein. The method is
XX CC useful for treating a patient suffering from a lysosomal storage disease
XX CC e.g. Fabry's disease. The present sequence represents the amino acid
XX CC sequence of soluble human GlcNAc-phosphotransferase.
XX SQ Sequence 1199 AA;
    Query Match 85.9%; Score 1487; DB 7; Length 1199;
    Best Local Similarity 100.0%; Pred. No. 2.9e-135;
    Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMPEEFDKTSFHKVRRHS 60
Db 919 DTFADSLRYVNKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMPEEFDKTSFHKVRRHS 978
QY 61 EDMQFAFSYFYLYMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db 979 EDMQFAFSYFYLYMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 1038
QY 121 GLEHMLNCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180
Db 1039 GLEHMLNCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 1098
QY 181 NKYFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDAQTVKAVL 240
Db 1099 NKYFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDAQTVKAVL 1158
QY 241 RDFYESMFPPIPSQFELPREYRNRFHMHLEQEWRAVDKDKL 281
Db 1159 RDFYESMFPPIPSQFELPREYRNRFHMHLEQEWRAVDKDKL 1199

RESULT 12
ABW01487
ID ABW01487 standard; protein; 1199 AA.
XX AC ABW01487;
XX XX 15-JAN-2004 (first entry)
XX DT
XX DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase.
XX KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
XX KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
XX KW gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;
XX KW kifunensine; glycosylation inhibition.
XX OS Unidentified.
XX OS US2003124652-A1.
XX PN
XX XX

```

PD 03-JUL-2003.
 XX 21-DEC-2001; 2001US-00023889.
 XX 21-DEC-2001; 2001US-00023889.
 XX (NOVA-) NOVAZYME PHARM INC.
 XX Canfield WM;
 XX WPI; 2003-810984/76.
 XX N-PSDB; AAD62490.
 XX Producing a high mannose glycoprotein for treating lysosomal storage
 PT disease, comprises culturing the lectin resistant mammalian cell in the
 PT presence of deoxymannojirimycin and kifunensine.
 XX
 XX Claim 8; Page 10-13; 46pp; English.
 XX
 XX The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
 CC (GlcNAc)-phosphotransferase
 XX
 XX Sequence 1199 AA;
 SQ
 Query Match 85.9%; Score 1487; DB 7; Length 1199;
 Best Local Similarity 100.0%; Pred. No. 2.9e-135;
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPPEFDKTSFHKVRS 60
 DB 919 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPPEFDKTSFHKVRS 978
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVDEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 120
 DB 979 EDMQFAFSFYLLMSAVQPLNISQVDEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 1038
 QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIKHAYKDK 180
 DB 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIKHAYKDK 1098
 QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHKAQTVKAVL 240
 DB 1099 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHKAQTVKAVL 1158
 QY 241 RDFSFMFPPIPSQPELPREYRNFLHMHLEQEWAYRDKDK 281
 DB 1159 RDFSFMFPPIPSQPELPREYRNFLHMHLEQEWAYRDKDK 1199
 RESULT 13
 ABW01536
 ID ABW01536 standard; protein; 1199 AA.
 XX
 XX AC ABW01536;
 XX
 XX DT 15-JAN-2004 (first entry)
 XX
 XX N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.
 XX
 XX Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 XX N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.
 XX
 XX Unidentified.
 XX OS
 XX

PN US2003124653-A1.
 XX 03-JUL-2003.
 XX 21-DEC-2001; 2001US-00023890.
 XX 21-DEC-2001; 2001US-00023890.
 XX (NOVA-) NOVAZYME PHARM INC.
 XX Canfield WM;
 XX WPI; 2003-810985/76.
 XX N-PSDB; AAD62649.
 XX Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 XX Claim 8; Page 10-13; 46pp; English.
 XX
 XX The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
 CC phosphotransferase protein
 XX
 XX Sequence 1199 AA;
 SQ
 Query Match 85.9%; Score 1487; DB 7; Length 1199;
 Best Local Similarity 100.0%; Pred. No. 2.9e-135;
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPPEFDKTSFHKVRS 60
 DB 919 DTFADSLRYVNKILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPPEFDKTSFHKVRS 978
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVDEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 120
 DB 979 EDMQFAFSFYLLMSAVQPLNISQVDEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 1038
 QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIKHAYKDK 180
 DB 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIKHAYKDK 1098
 QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHKAQTVKAVL 240
 DB 1099 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHKAQTVKAVL 1158
 QY 241 RDFSFMFPPIPSQPELPREYRNFLHMHLEQEWAYRDKDK 281
 DB 1159 RDFSFMFPPIPSQPELPREYRNFLHMHLEQEWAYRDKDK 1199
 RESULT 14
 ADG39858
 ID ADG39858 standard; protein; 248 AA.
 XX
 XX AC ADG39858;
 XX
 XX DT 26-FEB-2004 (first entry)
 XX
 XX Protein similar to human NOVI2 #4.
 XX
 XX Human; NOVI; cardiomyopathy; atherosclerosis; hypertension;
 KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;
 KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;
 KW congenital adrenal hyperplasia; prostate cancer; diabetes;
 KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;
 KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;
 KW multiple sclerosis; infectious disease; anorexia;
 KW

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; haematopoietic disorder; dyslipidaemia;
 KW wasting disorder.
 XX Homo sapiens.
 XX US2003203843-A1.
 XX 30-OCT-2003.
 XX
 FF 11-APR-2002; 2002US-00120801.
 XX
 XX 20-APR-2001; 2001US-0285609P.
 PR 23-APR-2001; 2001US-0285748P.
 PR 24-APR-2001; 2001US-0286068P.
 PR 25-APR-2001; 2001US-0286292P.
 PR 03-MAY-2001; 2001US-0288334P.
 PR 16-MAY-2001; 2001US-0291241P.
 PR 14-SEP-2001; 2001US-0322284P.
 XX
 XX (PENA/) PENA C E A.
 PA (GUOX/) GUO X.
 PA (SHIM/) SHIMKETS R A.
 PA (PADI/) PADIGARU M.
 PA (KEKU/) KEKUDA R.
 PA (SPYT/) SPYTEK K A.
 PA (MEHR/) MEHRABAN F.
 PA (TOPP/) TOPPER J N.
 PA (MALY/) MALYANKAR U M.
 PA (WASS/) WASSERMAN S M.
 PA (EDIN/) EDINGER S R.
 PA (SMIT/) SMITHSON G.
 PA (GUNT/) GUNTHER E.
 PA (KOMU/) KOMUVES L.
 XX
 XX PENA CEA, Guo X, Shimkets RA, Padigar M, Kekuda R, Spytek KA;
 PI Mehraban F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR,
 PI Smithson G, Gunther E, Komuves L;
 XX WPI: 2003-900671/82.
 DR
 XX
 XX New NOVX polypeptides and nucleic acids, useful for diagnosing or
 PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,
 PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or
 PT multiple sclerosis.
 XX
 XX Disclosure; SEQ ID NO 90; 215pp; English.
 XX
 XX The invention relates to a new isolated polypeptide comprising an amino
 CC acid sequence selected from 17 fully defined human NOVX sequences (even
 CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX
 CC amino acid or a variant of NOVX, where one or more amino acid residues in
 CC the variant differs in no more than 15% of the amino acid residues of
 CC NOVX. Also included are an isolated nucleic acid (NA) molecule
 CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above
 CC (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment
 CC encoding at least a portion of a NOVX polypeptide and a complement of
 CC NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an
 CC antibody that immunospecifically binds to NOVX, a method for determining
 CC the presence or amount of NOVX or NOVX NA in a sample, a method of
 CC identifying an agent that binds to NOVX, a method for identifying an
 CC agent that modulates the expression or activity of NOVX, a method for
 CC modulating the activity of NOVX, a method of treating or preventing a
 CC NOVX-associated disorder, a method for screening for a modulator of
 CC activity or of latency or predisposition to a NOVX-associated disorder, a
 CC method for determining the presence of or predisposition to a disease
 CC associated with altered levels of NOVX or NOVX NA in a first mammalian
 CC subject and a method of treating a pathological state in a mammal by
 CC administering NOVX or an antibody that binds to NOVX. The NOVX
 CC polypeptide, nucleic acid or antibody is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease or a
 CC NOVX-associated disorder. The NOVX polypeptides and nucleic acids
 CC encoding them are useful for diagnosing or treating pathologies, diseases

CC or conditions associated with NOVX sequences, including cardiomyopathy,
 CC atherosclerosis, hypertension, congenital heart defects, pulmonary
 CC stenosis, scleroderma, obesity, metabolic disturbances associated with
 CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal
 CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,
 CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,
 CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious
 CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,
 CC or Parkinson's disease), immune disorders, haematopoietic disorders,
 CC dyslipidaemias, and wasting disorders associated with chronic diseases.
 CC The polypeptides can be used as immunogens to produce antibodies and as
 CC vaccines. The sequences may further be used in chromosome mapping,
 CC identifying individual from minute biological samples (tissue typing),
 CC and in forensic identification of a biological sample. The present
 CC sequence is a protein showing sequence similarity to a NOVX protein.
 XX
 XX Sequence 248 AA;

Query Match 73.7%; Score 1276; DB 7; Length 248;
 Best Local Similarity 99.2%; Pred. No. 1.1e-115;
 Matches 242; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 85 VFDEVDTQSGVLSDRREIRTLATRIHELPLSLQDLTGLEHMLNCSKVLPAIDITQNNIP 144
 DB 5 VFDEVDTQSGVLSDRREIRTLATRIHELPLSLQDLTGLEHMLNCSKVLPAIDITQNNIP 64
 QY 145 PTQESYDNPVPPVTKSLVTNCKPVTDKIHKAYKDKNKYRFEIMGEEETAFKMIKRTNVSH 204
 DB 65 PTQESYDNPVPPVTKSLVTNCKPVTDKIHKAYKDKNKYRFEIMGEEETAFKMIKRTNVSH 124
 QY 205 VVGQLDDTRKQPRKPVCLNDINDHNHKAQVTKAVLRDFYFESMPFIPSQFELPRYRNRF 264
 DB 125 VVGQLDDTRKQPRKPVCLNDINDHNHKAQVTKAVLRDFYFESMPFIPSQFELPRYRNRF 184
 QY 265 LHMHELOQWRAVRDKLKEWTHCVLATLIMTFIPFPAEQIALKPKIPRRRIHKEASPN 324
 DB 185 LHMHELOQWRAVRDKLKEWTHCVLATLIMTFIPFPAEQIALKPKIPRRRIHKEASPN 244
 QY 325 RIRV 328
 DB 245 RIRV 248

RESULT 15
 ADG39855
 ID ADG39855 standard; protein; 663 AA.

XX AC ADG39855;
 XX DT 26-FEB-2004 (first entry)
 XX DE Protein similar to human NOV12 #1.
 XX Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;
 KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;
 KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;
 KW congenital adrenal hyperplasia; prostate cancer; diabetes;
 KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;
 KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;
 KW multiple sclerosis; infectious disease; anorexia;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; haematopoietic disorder; dyslipidaemia;
 KW wasting disorder.
 XX
 XX Homo sapiens.
 OS
 PN US2003203843-A1.
 XX
 XX 30-OCT-2003.
 PD
 XX
 XX 11-APR-2002; 2002US-00120801.
 PF
 XX 20-APR-2001; 2001US-0285609P.
 PR

PR 23-APR-2001; 2001US-0285748P.
PR 24-APR-2001; 2001US-0286068P.
PR 25-APR-2001; 2001US-0286292P.
PR 03-MAY-2001; 2001US-0288334P.
PR 16-MAY-2001; 2001US-0291241P.
PR 14-SEP-2001; 2001US-0322284P.
XX (PENA/) PENA C E A.
PA (GUOX/) GUO X.
PA (SHIM/) SHIMKETS R A.
PA (PADI/) PADIGARU M.
PA (KEKU/) KEKUDA R.
PA (SPYT/) SPYTEK K A.
PA (MEHR/) MEHRABAN F.
PA (TOPP/) TOPPER J N.
PA (WASS/) WASSERMAN U M.
PA (MALY/) MALYANKAR U M.
PA (WASS/) WASSERMAN S M.
PA (EDIN/) EDINGER S R.
PA (SMIT/) SMITHSON G.
PA (GUNT/) GUNTHER E.
PA (KOMU/) KOMUVES L.
XX
PI Pena CEA, Guo X, Shimkets RA, Padigaru M, Kekuda R, Spytek KA;
PI Mehrabab F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR;
PI Smithson G, Gunther E, Komuves L;
XX WPI; 2003-900671/82.
XX
XX New NOVX polypeptides and nucleic acids, useful for diagnosing or
PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,
PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or
PT multiple sclerosis.
XX
PS Disclosure; SEQ ID NO 87; 215pp; English.
XX
XX The invention relates to a new isolated polypeptide comprising an amino
CC acid sequence selected from 17 fully defined human NOVX sequences (even
CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX
CC amino acid or a variant of NOVX, where one or more amino acid residue in
CC the variant differs in no more than 15% of the amino acid residues of
CC NOVX. Also included are an isolated nucleic acid (NA) molecule
CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above
CC (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment
CC encoding at least a portion of a NOVX polypeptide and a complement of
CC NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an
CC antibody that immunospecifically binds to NOVX, a method for determining
CC the presence or amount of NOVX or NOVX NA in a sample, a method of
CC identifying an agent that binds to NOVX, a method for identifying an
CC agent that modulates the expression or activity of NOVX, a method for
CC modulating the activity of NOVX, a method of treating or preventing a
CC NOVX-associated disorder, a method for screening for a modulator of
CC activity or of latency or predisposition to a NOVX-associated disorder, a
CC method for determining the presence of or predisposition to a disease
CC associated with altered levels of NOVX or NOVX NA in a first mammalian
CC subject and a method of treating a pathological state in a mammal by
CC administering NOVX or an antibody that binds to NOVX. The NOVX
CC polypeptide, nucleic acid or antibody is useful in the manufacture of a
CC medicament for treating a syndrome associated with a human disease or a
CC NOVX-associated disorder. The NOVX polypeptides and nucleic acids
CC encoding them are useful for diagnosing or treating pathologies, diseases
CC or conditions associated with NOVX sequences, including cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, pulmonary
CC stenosis, scleroderma, obesity, metabolic disturbances associated with
CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal
CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,
CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,
CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious
CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,
CC or Parkinson's disease), immune disorders, haematopoietic disorders,
CC dyslipidaemias, and wasting disorders associated with chronic diseases.
CC The polypeptides can be used as immunogens to produce antibodies and as
CC vaccines. The sequences may further be used in chromosome mapping,
CC identifying individual from minute biological samples (tissue typing),

CC and in forensic identification of a biological sample. The present
CC sequence is a protein showing sequence similarity to a NOVX protein.
XX
SQ Sequence 663 AA;
Query Match 65.9%; Score 1140; DB 7; Length 663;
Best Local Similarity 98.6%; Pred. No. 8.3e-102; Indels 0; Gaps 0;
Matches 218; Conservative 1; Mismatches 2;
QY 1 DTAFDSLRYVKNKILNSKFGFTSRKVPAPHPMDRIVMQELQDMFPEEDKTSFHKVRHS 60
DB 413 DTAFDSLRYVKNKILNSKFGFTSRKVPAPHPMDRIVMQELQDMFPEEDKTSFHKVRHS 472
QY 61 EDMQFAFSFYVYLMMSAVOPLNISQVDFEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 120
DB 473 EDMQFAFSFYVYLMMSAVOPLNISQVDFEVDTDOSGVLSDREIRTLATRIHELPLSLQDLT 532
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTNCKPVTDKTHKAYKDK 180
DB 533 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTNCKPVTDKTHKAYKDK 592
QY 181 NKYRFEIMGEEETAFKMRITNVSHVVGQLDDIRKNRKFVC 221
DB 593 NKYRFEIMGEEETAFKMRITNVSHVVGQLDDIRKNRPRISLC 633

Search completed: November 21, 2004, 12:54:46

Job time : 39.6766 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:47:18 ; Search time 9.88676 Seconds
(without alignments)
2200.144 Million cell updates/sec

Title: US-10-023-888-5
Perfect score: 1731
Sequence: 1 DTFADSLRYVVKILNSKFGF.....RKIFPRRIHKEASPNRIRV 328

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	US-09-635-872A-2
2	1731	100.0	328	4	US-09-636-077A-2
3	1731	100.0	328	4	US-09-636-060C-2
4	1731	100.0	328	4	US-09-986-552-2
5	1731	100.0	328	4	US-09-636-596C-2
6	1698	98.1	328	4	US-09-635-872A-8
7	1698	98.1	328	4	US-09-636-077A-8
8	1698	98.1	328	4	US-09-636-060C-8
9	1698	98.1	328	4	US-09-986-552-8
10	1698	98.1	328	4	US-09-636-596C-8
11	402	23.2	502	4	US-09-635-872A-13
12	402	23.2	502	4	US-09-636-077A-13
13	402	23.2	502	4	US-09-636-060C-13
14	402	23.2	502	4	US-09-986-552-13
15	402	23.2	502	4	US-09-636-596C-13
16	269	15.5	299	4	US-09-270-767-37559
17	269	15.5	299	4	US-09-270-767-52776
18	145	8.4	28	4	US-09-635-872A-27
19	145	8.4	28	4	US-09-636-077A-27
20	145	8.4	28	4	US-09-636-060C-27
21	145	8.4	28	4	US-09-986-552-27
22	145	8.4	28	4	US-09-636-596C-27
23	116	6.7	44	4	US-09-270-767-33278
24	116	6.7	44	4	US-09-270-767-48495
25	113	5.6	545	4	US-08-926-107A-10
26	101	5.8	2662	4	US-09-595-684B-31
27	101	5.8	2662	4	US-09-538-092-1252

RESULT 2

ALIGNMENTS

RESULT 1
US-09-635-872A-2
; Sequence 2, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635, 872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-635-872A-2

Query Match	100.0%;	Score 1731;	DB 4;	Length 328;
Best Local Similarity	100.0%;	Pred. No. 1.5e-168;		
Matches	328;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	DTFADSLRYVVKILNSKFGFTSRKVPAPHPMHMIDRVMOELQDMFPPEFDKTSFHKVRHS	60	
Db	1	DTFADSLRYVVKILNSKFGFTSRKVPAPHPMHMIDRVMOELQDMFPPEFDKTSFHKVRHS	60	
QY	61	EDMQPAPSYFYLLMSAVQPLNISQVDFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT	120	
Db	61	EDMQPAPSYFYLLMSAVQPLNISQVDFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT	120	
QY	121	GLEHMLNCSKMLPADITQLANNIPPTQESYVDNLPVTKSLVNTCKEVTDKIHKAYKDK	180	
Db	121	GLEHMLNCSKMLPADITQLANNIPPTQESYVDNLPVTKSLVNTCKEVTDKIHKAYKDK	180	
QY	181	NKYRFEIINGEIEEIAFMIRTNVSHVVGQDDIRKNPKRFVCLNDNIDHNHDKDAQVTKAVL	240	
Db	181	NKYRFEIINGEIEEIAFMIRTNVSHVVGQDDIRKNPKRFVCLNDNIDHNHDKDAQVTKAVL	240	
QY	241	RDVFSMPPIPSQFELPREYRNRLHMEQEWAYRDKLFWTHCVLATLIMFTIPSF	300	
Db	241	RDVFSMPPIPSQFELPREYRNRLHMEQEWAYRDKLFWTHCVLATLIMFTIPSF	300	
QY	301	AEQLIALKKRPIPRRIHKEASPNRIRV	328	
Db	301	AEQLIALKKRPIPRRIHKEASPNRIRV	328	

US-09-636-077A-2		US-09-636-077A-2	
; Sequence 2, Application US/09636077A		; Sequence 2, Application US/09636077A	
; Patent No. 6537785		; Patent No. 6537785	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	
; APPLICANT: CANFIELD, WILLIAM		; APPLICANT: CANFIELD, WILLIAM	
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE		; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE	
; FILE REFERENCE: 195612US0		; FILE REFERENCE: 195612US0	
; CURRENT APPLICATION NUMBER: US/09/636, 077A		; CURRENT APPLICATION NUMBER: US/09/636, 077A	
; CURRENT FILING DATE: 2000-08-10		; CURRENT FILING DATE: 2000-08-10	
; PRIOR APPLICATION NUMBER: 60/153,831		; PRIOR APPLICATION NUMBER: 60/153,831	
; PRIOR FILING DATE: 1999-09-14		; PRIOR FILING DATE: 1999-09-14	
; NUMBER OF SEQ ID NOS: 52		; NUMBER OF SEQ ID NOS: 52	
; SOFTWARE: Patentin version 3.0		; SOFTWARE: Patentin version 3.1	
; SEQ ID NO 2		; SEQ ID NO 2	
; LENGTH: 328		; LENGTH: 328	
; TYPE: PRT		; TYPE: PRT	
; ORGANISM: Homo sapiens		; ORGANISM: Homo sapiens	
US-09-636-077A-2		US-09-636-060C-2	
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
DB	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
RESULT 4		RESULT 3	
US-09-986-552-2		US-09-636-060C-2	
; Sequence 2, Application US/09986552		; Sequence 2, Application US/09636060C	
; Patent No. 6670165		; Patent No. 6642038	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	
; APPLICANT: CANFIELD, WILLIAM		; APPLICANT: CANFIELD, WILLIAM M	
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES		; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY	
; FILE REFERENCE: 215089US77DIV		; FILE REFERENCE: 210119US0CONT	
; CURRENT APPLICATION NUMBER: US/09/986,552		; CURRENT APPLICATION NUMBER: US/09/636,060C	
; CURRENT FILING DATE: 2001-11-09		; CURRENT FILING DATE: 2000-08-10	
; PRIOR APPLICATION NUMBER: 09/635,872		; PRIOR APPLICATION NUMBER: 60/153,831	
; PRIOR FILING DATE: 2000-08-10		; PRIOR FILING DATE: 1999-09-14	
; PRIOR APPLICATION NUMBER: 60/153,831		; NUMBER OF SEQ ID NOS: 52	
; PRIOR FILING DATE: 1999-09-14		; SOFTWARE: Patentin version 3.1	
; NUMBER OF SEQ ID NOS: 52		; SEQ ID NO 2	
; SOFTWARE: Patentin version 3.1		; LENGTH: 328	
; SEQ ID NO 2		; TYPE: PRT	
; LENGTH: 328		; ORGANISM: Homo sapiens	
; TYPE: PRT		; ORGANISM: Homo sapiens	
; ORGANISM: Homo sapiens		US-09-636-060C-2	
US-09-986-552-2		Query Match	
Query Match		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Best Local Similarity 100.0%; Pred. No. 1.5e-168;		Best Local Similarity 100.0%; Pred. No. 1.5e-168;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
DB	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
DB	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVYKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPRKFVCLNDNDHNNHKAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328
DB	301	AEQIALKRRKIFPRRRIHKEASPNRIY	328

```
RESULT 5
US-09-636-596C-2
; Sequence 2, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 19929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636.596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-596C-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.5e-168;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Db 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328

RESULT 6
US-09-635-872A-8
; Sequence 8, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES
; FILE REFERENCE: 195613USO
; CURRENT APPLICATION NUMBER: US/09/635.872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Db 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328

RESULT 7
US-09-636-077A-8
; Sequence 8, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612USO
; CURRENT APPLICATION NUMBER: US/09/636.077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNKILNSKFGTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Db 241 RQFYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKEWTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRRIHKEASPNRIRV 328
```

Db 301 AEQIIALKRKIFPRRRHKASPDRIYV 328

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
US-09-636-060C-8
; Sequence 8, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240

QY 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
Db 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

QY 301 AEQIIALKRKIFPRRRHKASPDRIYV 328
Db 301 AEQIIALKRKIFPRRRHKASPDRIYV 328

RESULT 9
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240

QY 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
Db 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

QY 301 AEQIIALKRKIFPRRRHKASPDRIYV 328
Db 301 AEQIIALKRKIFPRRRHKASPDRIYV 328

RESULT 9
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240

QY 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
Db 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

RESULT 10
US-09-636-596C-8
; Sequence 8, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLCNACASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-596C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYLLMSAVQPLNISQVFHEVDTDSQVLSDRIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESVYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEBEEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL 240

QY 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300
Db 241 RDFYESMFPPIPSQPELPREYRNRLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300


```

; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-986-552-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.8e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

Qy 1 DTFADSLRYNKILNSKFGTSRKVPAPMHPMDIRVMQELQDMFPEEPDKTSFHKVHS 60
Db 259 DIYSHSLIATNMLNRYAGFKARHVLAHVGFLDKDIVEAMQRRFHQOILDTAHQRFAP 318
Qy 61 EDMQFARSFYFLMSAVQPLNISQVDFEVDVTDOSGVLSDREIRTLATRIHELPLSLQDLT 120
Db 319 TDLQYAFAYISFLMSETKWMSVEIFDFEFDGTSATWSDEVRTFLTRIYQPPLDWSAMR 378
Qy 121 GLEHMLINCSK---MLPADITQLNIPPTQESYYDPNLPVTKSLVTNCKPVTDKIHKA 176
Db 379 YFEVWQNCNRLNGLMHLKVDVTEHSTL--VYERYEDSNLPTITRDLVVRCPPLAEALAN 436
Qy 177 YKDKNKYRFEIMG- -BEIAFKMIRTNVSHVVGOLDDIRKNPKFVCLNDNIDHN 229
Db 437 FAVRPKNFHVSPKRTSHSNFMMLTSLNTEVVESLDRLRNPKFNCINDNLDAN 491

RESULT 15
US-09-636-596C-13
; Sequence 13 Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLCNase OF THE LYXOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-636-596C-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.8e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

Qy 1 DTFADSLRYNKILNSKFGTSRKVPAPMHPMDIRVMQELQDMFPEEPDKTSFHKVHS 60
Db 259 DIYSHSLIATNMLNRYAGFKARHVLAHVGFLDKDIVEAMQRRFHQOILDTAHQRFAP 318
Qy 61 EDMQFARSFYFLMSAVQPLNISQVDFEVDVTDOSGVLSDREIRTLATRIHELPLSLQDLT 120
Db 319 TDLQYAFAYISFLMSETKWMSVEIFDFEFDGTSATWSDEVRTFLTRIYQPPLDWSAMR 378
Qy 121 GLEHMLINCSK---MLPADITQLNIPPTQESYYDPNLPVTKSLVTNCKPVTDKIHKA 176
Db 379 YFEVWQNCNRLNGLMHLKVDVTEHSTL--VYERYEDSNLPTITRDLVVRCPPLAEALAN 436
Qy 177 YKDKNKYRFEIMG- -BEIAFKMIRTNVSHVVGOLDDIRKNPKFVCLNDNIDHN 229
Db 437 FAVRPKNFHVSPKRTSHSNFMMLTSLNTEVVESLDRLRNPKFNCINDNLDAN 491

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:56:09 ; Search time 35.4053 Seconds
(without alignments)
3280.691 Million cell updates/sec

Title: US-10-023-888-5
Perfect score: 1731
Sequence: 1 DTFADSLRYVVKILNSKFGF.....RKIPFRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pap.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pap.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pap.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	328	9	US-09-895-072-2
2	1731	100.0	328	9	US-09-986-552-2
3	1731	100.0	328	14	US-10-023-888-5
4	1731	100.0	328	14	US-10-023-889-5
5	1731	100.0	328	14	US-10-023-890-5
6	1731	100.0	328	14	US-10-024-197-5
7	1731	100.0	328	14	US-10-023-894-5
8	1731	100.0	328	14	US-10-306-686-2
9	1698	98.1	328	9	US-09-895-072-8
10	1698	98.1	328	9	US-09-986-552-8
11	1698	98.1	328	14	US-10-023-888-10
12	1698	98.1	328	14	US-10-023-889-10
13	1698	98.1	328	14	US-10-023-890-10

14	1698	98.1	328	14	US-10-024-197-10	Sequence 10, Appl
15	1698	98.1	328	14	US-10-023-894-10	Sequence 10, Appl
16	1698	98.1	328	14	US-10-306-686-8	Sequence 8, Appl
17	1487	85.9	1199	14	US-10-023-888-2	Sequence 2, Appl
18	1487	85.9	1199	14	US-10-023-889-2	Sequence 2, Appl
19	1487	85.9	1199	14	US-10-023-890-2	Sequence 2, Appl
20	1487	85.9	1199	14	US-10-024-197-2	Sequence 2, Appl
21	1487	85.9	1199	14	US-10-023-894-2	Sequence 2, Appl
22	1290	74.5	248	14	US-10-120-801-90	Sequence 90, Appl
23	1140	65.9	663	14	US-10-120-801-87	Sequence 87, Appl
24	1140	65.9	1459	14	US-10-120-801-26	Sequence 26, Appl
25	581	33.6	384	14	US-10-120-801-89	Sequence 89, Appl
26	506	29.2	652	14	US-10-120-801-91	Sequence 91, Appl
27	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appl
28	402	23.2	502	9	US-09-986-552-13	Sequence 13, Appl
29	402	23.2	502	14	US-10-023-888-16	Sequence 16, Appl
30	402	23.2	502	14	US-10-023-889-16	Sequence 16, Appl
31	402	23.2	502	14	US-10-023-890-16	Sequence 16, Appl
32	402	23.2	502	14	US-10-024-197-16	Sequence 16, Appl
33	402	23.2	502	14	US-10-023-894-16	Sequence 16, Appl
34	402	23.2	502	14	US-10-306-686-13	Sequence 13, Appl
35	300	17.3	55	14	US-10-029-386-30775	Sequence 30775, A
36	228	13.2	112	13	US-10-079-623-349	Sequence 349, Appl
37	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appl
38	145	8.4	28	9	US-09-986-552-27	Sequence 27, Appl
39	145	8.4	28	14	US-10-306-686-27	Sequence 27, Appl
40	124	7.2	1312	14	US-10-369-493-1950	Sequence 1950, Ap
41	112	6.5	46	14	US-10-023-888-26	Sequence 26, Appl
42	110.5	6.4	1015	16	US-10-437-963-160369	Sequence 160369,
43	104.5	6.0	937	16	US-10-437-963-156445	Sequence 156445,
44	104.5	6.0	1999	14	US-10-028-248A-107	Sequence 107, App
45	104.5	6.0	1999	15	US-10-107-782-107	Sequence 107, App

ALIGNMENTS

RESULT 1
US-09-895-072-2
; Sequence 2, Application US/09895072
; Patent No. US2002025550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119050CONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-072-2

Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DTFADSLRYVVKILNSKFGTISRKYPAHMPHMDRIVMQELODMPEBFDKTSFKVRRHS	60
DB	1	DTFADSLRYVVKILNSKFGTISRKYPAHMPHMDRIVMQELODMPEBFDKTSFKVRRHS	60
QY	61	EDMQAFSFFYYLMSAVQPLNISQVFDEVDITDQSGVLSDRERTLATRIHELPLSLQDIT	120
DB	61	EDMQAFSFFYYLMSAVQPLNISQVFDEVDITDQSGVLSDRERTLATRIHELPLSLQDIT	120
QY	121	GLEHMLNCSKMLPADITQLNIPQTSYDNPVPTKSLVTNCKVPTDKIHAYKDK	180

Db 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Qy 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Db 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Qy 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
RESULT 2
US-09-986-552-2
; Sequence 2, Application US/05986552
; Patent No. US20020150381A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-2
Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Qy 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Db 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Qy 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Db 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Qy 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
RESULT 3
US-10-023-888-5
; Sequence 5, Application US/10023889
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-5
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Qy 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Db 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180

; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-5
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Qy 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Db 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Qy 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Db 181 NKYRFEIMGEEIAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240
Qy 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYEMFPIPSQFELPREYRNFLMHQELQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328
RESULT 4
US-10-023-889-5
; Sequence 5, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-889-5
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKYRHS 60
Qy 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFLMSAVQPLNISQVDFEVDTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180
Db 121 GLEHMLNCSKMLPADITQLNN:PPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180

Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

RESULT 5

US-10-023-890-5
; Sequence 5, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYDRATE
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

RESULT 6

US-10-024-197-5
; Sequence 5, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREOSIDASE AND METHODS OF
; FILE REFERENCE: 209794USO
; CURRENT APPLICATION NUMBER: US/10/024,197

; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Db 181 NKYRFEIMGEETAFKMIKTNTVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

RESULT 7

US-10-023-894-5
; Sequence 5, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFIELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-149;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 DB 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 QY 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 DB 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 QY 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328

RESULT 8

US-10-306-686-2
 ; Sequence 2, Application US/10306686
 ; Publication No. US20030148460A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY
 ; FILE REFERENCE: 230397US77DIV
 ; CURRENT APPLICATION NUMBER: US/10/306,686
 ; CURRENT FILING DATE: 2002-11-29
 ; PRIOR APPLICATION NUMBER: 09/636,596
 ; PRIOR FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-08-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 328
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-306-686-2

Query Match 100.0%; Score 1731; DB 14; Length 328;
 Best Local Similarity 100.0%; Pred. No. 4.3e-149;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60
 QY 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVTKSLVNTCKPVTDKIKHAYKOK 180
 DB 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVTKSLVNTCKPVTDKIKHAYKOK 180
 QY 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 DB 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 QY 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 DB 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 QY 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328

RESULT 9

US-09-895-072-8
 ; Sequence 8, Application US/09895072
 ; Patent No. US2002025550A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM M
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 ; FILE REFERENCE: 210119SUOCONT

; CURRENT APPLICATION NUMBER: US/09/895,072
 ; CURRENT FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; PRIOR APPLICATION NUMBER: US 09/635,872
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 328
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-895-072-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
 Best Local Similarity 97.6%; Pred. No. 4.4e-146;
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60
 QY 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 DB 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVTKSLVNTCKPVTDKIKHAYKOK 180
 DB 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVTKSLVNTCKPVTDKIKHAYKOK 180
 QY 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 DB 181 NKYFEIMGEETAFKIRTNVSHVVGOLDIRKNPKFVCLNDNDHNDKDAQTAVL 240
 QY 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 DB 241 RDFSMEFPISQPELPREYRNFLHMHQEWAYRDKLFWTHCVLATLIMFTISFF 300
 QY 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328

RESULT 10

US-09-986-552-8
 ; Sequence 8, Application US/09986552
 ; Patent No. US20020150981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 ; FILE REFERENCE: 215089US77DIV
 ; CURRENT APPLICATION NUMBER: US/09/986,552
 ; CURRENT FILING DATE: 2001-11-09
 ; PRIOR APPLICATION NUMBER: 09/635,872
 ; PRIOR FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 328
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-986-552-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
 Best Local Similarity 97.6%; Pred. No. 4.4e-146;
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMIDRIYVQELQDMFPEEPDKTSFHKVRS 60

QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHDLPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPTQEAAYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

RESULT 11

US-10-023-888-10
; Sequence 10, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-888-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 4.4e-146;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
DB 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHDLPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPTQEAAYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

RESULT 12

US-10-023-889-10
; Sequence 10, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-889-10

Query Match 98.1%; Score 1698; DB 14; Length 328;

Best Local Similarity 97.6%; Pred. No. 4.4e-146;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
DB 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHDLPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPTQEAAYDPNLPVTKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDCLKFWTHCVLATLIMFTIPSPFF 300
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

RESULT 13

US-10-023-890-10
; Sequence 10, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-890-10

Query Match 98.1%; Score 1698; DB 14; Length 328;

Best Local Similarity 97.6%; Pred. No. 4.4e-146;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
DB 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHPMHMIDRIVMQELQDMFPPEFDKTSFHKVVRHS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHDLPLSLQDLT 120

; TITLE OF INVENTION: EXPRESSION OF LYOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: ACETYLGUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDAS
; CURRENT APPLICATION NUMBER: US/10/023.894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-023-894-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 4.4e-146;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMHIDRIVMQELQDMFPEEFDKTSFHKVYRHS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMHIDRIVMQELQDMFPEEFDKTSFHKVYRHS 60
QY 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHDLPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
DB 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
QY 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
DB 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
QY 301 AEQIIALKRKIFPRRRHKEASPNRIY 328
DB 301 AEQIIALKRKIFPRRRHKEASPNRIY 328

Search completed: November 21, 2004, 13:22:35
Job time : 36.4053 secs

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
DB 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
QY 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
DB 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
QY 301 AEQIIALKRKIFPRRRHKEASPNRIY 328
DB 301 AEQIIALKRKIFPRRRHKEASPNRIY 328

RESULT 14
US-10-024-197-10
; Sequence 10, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBOSEDASE AND METHODS
; FILE REFERENCE: 209794USO
; CURRENT APPLICATION NUMBER: US/10/024.197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-024-197-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 4.4e-146;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMHIDRIVMQELQDMFPEEFDKTSFHKVYRHS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMHIDRIVMQELQDMFPEEFDKTSFHKVYRHS 60
QY 61 EDMQFAFSFYFLMSAVQPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHDLPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTNCKPVTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
DB 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTAVKL 240
QY 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
DB 241 RDFYESMFPIPSQPELPREYRNRLHMHQELQEWAYRDKLFWTHCVLATLIMFTIFSFF 300
QY 301 AEQIIALKRKIFPRRRHKEASPNRIY 328
DB 301 AEQIIALKRKIFPRRRHKEASPNRIY 328

RESULT 15
US-10-023-894-10
; Sequence 10, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFIELD, Stuart

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:39:03 ; Search time 10.0204 Seconds
(without alignments)
3149.495 Million cell updates/sec

Title: US-10-023-888-5

Perfect score: 1731

Sequence: 1 DTFADSLRYNKLNSKPGF.....RKIPFRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290	74.5	248	2 T50618	hypothetical prote
2	581	33.6	384	2 I49528	hypothetical prote
3	124	7.2	1312	1 BMYDL	RAD50 protein - ye
4	113.5	6.6	545	2 C82014	probable capsula b
5	111	6.4	3119	2 T18414	protein g377 - mal
6	106.5	6.2	971	2 T00394	GTP-binding protei
7	104.5	6.0	1999	1 S21801	myosin heavy chain
8	103	6.0	690	2 P71239	hypothetical prote
9	102.5	5.9	706	2 D90124	hypothetical prote
10	102	5.9	515	2 S59811	probable segregati
11	101.5	5.9	301	2 AH1144	vacuolar membrane
12	101	5.8	2663	1 S28261	centromere protein
13	99.5	5.7	587	2 S20596	immediate early pr
14	99	5.7	1961	1 A61231	myosin heavy chain
15	97.5	5.6	735	2 D70174	methy/-accepting c
16	97.5	5.6	1498	2 AG1439	B. subtilis Yuka p
17	97	5.6	586	2 T20009	hypothetical prote
18	96.5	5.6	574	2 T33794	hypothetical prote
19	96	5.5	586	2 T34703	hypothetical prote
20	95.5	5.5	1837	2 T41023	probable nuclear p
21	95	5.5	651	2 F89798	hypothetical prote
22	95	5.5	1252	2 B42771	reticulocyte-bindl
23	94	5.4	718	2 G97249	transcription acce
24	94	5.4	810	2 S49744	AMP deaminase (EC
25	94	5.4	817	2 S53921	hypothetical prote
26	94	5.4	1979	1 S03166	myosin heavy chain
27	93.5	5.4	353	2 E70100	tryptophan-tRNA li
28	93.5	5.4	1101	2 AB2481	hypothetical prote
29	93	5.4	314	2 A71614	syntaxin PFE0480W

conserved hypothet
hypothetical prote
apolipoprotein A-1
hypothetical prote
lantibiotic mersac
probable multiple
hypothetical prote
H+-exporting ATPas
unknown protein. 7
DNA-directed DNA p
DNA-directed DNA p
hypothetical prote
peptide synthetase
dTDP-glucose 4-6-d
trans-activating t
early gene transac

ALIGNMENTS

RESULT 1

T50618

hypothetical protein DKFZp762B226.1 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004

C;Accession: T50618

R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, June 2000

A;Reference number: Z25143

A;Accession: T50618

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-248 <AAA>

A;Cross-references: UNIPROT:Q9NPW9; EMBL:AL359588

A;Experimental source: adult melanoma (MeWo cell line); clone DKFZp762B226

C;Genetics:

A;Note: DKFZp762B226.1

Query Match 74.5%; Score 1290; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 7.6e-87;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VFDEVDTDGSLDREIRITLRIHEPLSLQDLTGLEHMLNCKMLPADITQLNNIP 144
Db 5 VFDEVDTDGSLDREIRITLRIHEPLSLQDLTGLEHMLNCKMLPADITQLNNIP 64

QY 145 PTQESYDNLPPVTKSLVTNCKPVTDKIHKAYKDKNKYRFEIMGEEELAFKMRTNVSH 204
Db 65 PTQESYDNLPPVTKSLVTNCKPVTDKIHKAYKDKNKYRFEIMGEEELAFKMRTNVSH 124

QY 205 VVGQLDDIRKNPRKFCVCLNDNDHDKDAQTVAVLRFYFESMFPPIPSQFELPREYRNF 264
Db 125 VVGQLDDIRKNPRKFCVCLNDNDHDKDAQTVAVLRFYFESMFPPIPSQFELPREYRNF 184

QY 265 LHMHELOEQRAYRDKLKFVTHCVLATLIMFTIFSPFAEGLIALKRIPRRRIHKEASN 324
Db 185 LHMHELOEQRAYRDKLKFVTHCVLATLIMFTIFSPFAEGLIALKRIPRRRIHKEASN 244

QY 325 RIRV 328
Db 245 RIRV 248

RESULT 2

I49528

hypothetical protein - mouse (fragment)

N;Alternate names: cDNA1

C;Species: Mus musculus (house mouse)

C;Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004

C;Accession: I49528

R;Cordes, S.P.; Barsh, G.S.

Cell 79, 1025-1034, 1994

A:Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transmembrane protein
A:Reference number: A55200; MUID:95094266; PMID:8001130

A:Accession: T49528

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-384 <RES>

A:Cross-references: UNIPROT:Q61340; GB:L36434; NID:G625041; PIDN:AAA65688.1; PID:G625042

Query Match 33.6%; Score 581; DB 2; Length 384;

Best Local Similarity 97.4%; Pred. No. 5.9e-35;

Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYNKNILNSKFGTSSKVPFAHMPHMDIRVMOELQDMFPEEDKTSFHKVRHS 60

Db 267 DTFADSLRYNKNILNSKFGTSSKVPFAHMPHMDIRVMOELQDMFPEEDKTSFHKVRHS 326

Qy 61 EDWQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLS 115

Db 327 EDWQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHDPLT 381

RESULT 3

BWBYDL

RAD50 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein N0872; protein YNL250W

C:Species: Saccharomyces cerevisiae

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004

A:Accession: S05808; S63223

R:Alani, E.; Subbiah, S.; Kleckner, N.

Genetics 122, 47-57, 1989

A:Title: The yeast RAD50 gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain

A:Reference number: S05808; MUID:89276917; PMID:2659437

A:Accession: S05808

A:Molecule type: DNA

A:Residues: 1-1312 <ALA>

A:Cross-references: UNIPROT:P12753; EMBL:X14814; NID:G4272; PIDN:CAA32919.1; PID:G4273

R:Sen-Gupta, M.; Guelidener, U.; Beinbauer, J.; Fiedler, T.; Hegemann, J.H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63220

A:Accession: S63223

A:Molecule type: DNA

A:Residues: 1-1312 <SEN>

A:Cross-references: EMBL:Z71526; NID:G1302292; PIDN:CAA96157.1; PID:G1302293; GSPDB:GN00

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:RAD50; MIPS:YNL250W

A:Cross-references: SGD:S0005194; MIPS:YNL250W

A:Map position: 14L

C:Superfamily: RAD50 protein

C:Keywords: ATP; coiled coil; DNA repair; meiosis; nucleus

F:177-421/Region: heptad repeats

F:743-995/Region: heptad repeats

F:407/Binding site: Atp (lys) #status predicted

Query Match

Best Local Similarity 7.2%; Score 124; DB 1; Length 1312;

Matches 68; Conservative 65; Mismatches 126; Indels 60; Gaps 14;

Qy 5 DLSRYNKNILNSKFGTSSKVPFAHMPHMDIRVMOELQDMFPEEDKTSF 54

Db 189 DNLKSIKXDMVDVTKLQKQSV-EHLKLDKDSKAMKLNHQLQTKIDYNEVESEISQL 247

Qy 55 HKVRHSDMQAFASY-FYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPL 113

Db 248 NEITEKSKLFSKQDFKILSKVENLKNVTKL-----SISD-QVKRLNSIDILD 296

Qy 114 LSLQDLTGLEHMLNCSKMLPADITOLNNIPTQESYYD--PNLPPVTKSLVNTCKPVT 171

Db 297 LSKPD---LQNLNANFSKVLMDKXNQSLDLETTISLKDQSSQLSLSNLRQGL-- 351

Qy 172 KHKAYDKNRYFRIMGEIEIAFKVIRTVSHV-----VGQLDDIRKNPKFVC--LNDN 225

Db 352 EAGKETYEKNRHLSSLKX---AFQHKFQGLSNIENSMDAQVNHEMSQKRAFISQDLTDT 408

Qy 226 IDNHKDAQTVKAVLRDIFYSMFPPIPSQFELPREYRNFHL----- 266

Db 409 IDQFAKDITQLKETNLSDLIKSITVDSONLEYNKDRSKLIHDSLEAEKLSKFKSLSTQD 468

Qy 267 --MHELQEWAYRDKLKFW 283

Db 469 SLNHELENLTKYKELQSW 487

RESULT 4

CS2014

probable capsule biosynthesis protein NMA0200 [imported] - Neisseria meningitidis (strain

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004

A:Accession: CS2014

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:2022556; PMID:10761919

A:Accession: CS2014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <PAR>

A:Cross-references: UNIPROT:Q9JWW8; GB:AL162752; GB:AL157959; NID:G7378778; PIDN:CA8835

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: sacB; NMA0200

Query Match

Best Local Similarity 6.6%; Score 113.5; DB 2; Length 545;

Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Qy 3 FADSLRYNKNILNSKFGTSSKVPFAHMPHMDIRVMOELQDMFPEEDKTSFHKVRHSD 62

Db 395 YLNGARNANTLLEKFKFTKLTHTSPQSWRTDILFEMEKYFEENRNLHNFRLSD 454

Qy 63 MQF-AFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQLTG 121

Db 455 IAVTGYLYHH-----ALLSGLALQS-----SDKTE 480

Qy 122 LEHMLNCSKMLPADITOLNNIPTQESYVDNLPVTKSLVNTCKPVTDKHKAYKDK 180

Db 481 LVQQNHDPKK-----KLNNVVTLTKERNFD-KLP----- 508

Qy 181 NKRYFEIMGEBEIAFAPKIRNTVSHVVGQLDDIRKNPKFVCLNDNIDHNNKDAQTVKAVL 240

Db 509 -----LSVCINDGADSHLNEWNVQVI- 530

Qy 241 RDPYESMFPPIPSQFE 255

Db 531 -KFELETLPPLSPSFE 544

RESULT 5

T18414

protein g377 - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

A:Accession: T18414

R:Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Cappai, R.

Mol. Biochem. Parasitol. 74, 143-156, 1995

A:Title: The Leishmania promastigote surface antigen 2 complex is differentially expressed

A:Reference number: Z18933; MUID:96360472; PMID:8719156

A:Accession: T18414

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3119 <HAN>

A:Cross-references: UNIPROT:Q25857; EMBL:L04161; NID:G309687; PID:G309688; PIDN:AAAC3725

Query Match

Best Local Similarity 6.4%; Score 111; DB 2; Length 3119;

Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LQDMFFPEEDFTSFHKVHSEDMQFAFSFYFYLLMSAVQPLNISQVDEVDTDQSGVLSLSD- 99
 Db 1261 LEDIWBE---KKKYKEEYSKWRRIISNLDYEVNK---QIKHHYHKVDI-----ISEH 1308
 QY 100 --REIR-----TLATRIHELPLSLQ-----DLTGLEHMLNCSKMLPADITQLNIPP-- 145
 Db 1309 KFQEIQRHMRDKIENTHELYKEMYVQIQIDLTNYHQLENIHSELLQALQOQNKNI PRHL 1368
 QY 146 -----TQESYVDNPLPV--TKSLVTNCKPVTDKLHKAYKD-----KKYKRFEL--- 187
 Db 1369 NVLEKKLEIITKK--KKVKPDITSSHATDEQQVSDTLIRGAHHGDIILKGEDNDEVLLI 1426
 QY 188 -----MGEEIEAFKMIRTNVSHVVQG-----LDDIRKNPRKPVCLINDNIDH 228
 Db 1427 EQIQSLKTKMGDNQNGVGSILEKLNLSLSDQYQLLQDKLVNVEDIYKNLRF---KHVIEK 1483
 QY 229 NHKDAQ-----TVKAVLRDFEYSMPFIPSQFELPREYRNRFELMHLEQWEAYRDKLK 281
 Db 1484 LHKESKINREKFTIKVDVLSNVYSTL-----EYWKVFL-LHDFQEWSEFKDELE 1531

RESULT 6

T00394
 C: Species: cherry salmon
 C: Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
 C: Accession: T00394
 R: Kubokawa, K.; Miyashita, T.; Kubo, Y.
 FEBS Lett. 431, 231-235, 1998
 A: Title: Isolation of a cDNA for a novel 120-kDa GTP-binding protein expressed in motor
 A: Reference number: Z14145; MUID: 98372640; PMID: 9708909
 A: Accession: T00394
 A: Status: preliminary; translated from GB/EMBL/DBD
 A: Molecule type: mRNA
 A: Residues: 1-971 <KUB>
 A: Cross-references: UNIPROT:O93248; EMBL:AB012720; NID:g3413519; PID:g3413519; PID:g3413519
 C: Gene: MG120
 C: Keywords: GTP binding; nucleotide binding; P-loop
 F: 307-314/Region: nucleotide-binding motif A (P-loop)

Query Match 6.2%; Score 106.5; DB 2; Length 971;
 Best Local Similarity 23.0%; Pred. No. 6.9;
 Matches 62; Conservative 43; Mismatches 79; Indels 85; Gaps 14;

QY 7 LRYUNKILNSKF--GFTSRKVPAPHPMIDRIVWQELQDMFPPEFDKTSFKVHSEDMQ 64
 Db 102 LRYI--VLGSAVGGYTAKK-----TYEWKMLPD----- 130
 QY 65 FAFSYFYLLMSA---VOPLNISQVDEVDTDQ--SGVLSDRREIRTLATRIHELPLSLQDLT 120
 Db 131 -----NSAYNWIPDFVWLSQDIDLDKTLKP--ELEEIAKLPLP-----DF- 174
 QY 121 GLEHMLNCSKMLPADITQLNIPPQTESYVDNPLPVTKSLVTNCKPVTDKTHKAYKDK 180
 Db 175 -----DKIGENFTFLKSILFTAEPAGDPVKAATEAPVTATPEASDKQFKSSDK 224
 QY 181 NKYRFEIMGEEIEAFKMIRTNVSH--VVGQLDDIRKNPRKPVCLINDNID--HNHKDAQTVK 237
 Db 225 EKV-----DQIQEELLRTQMKYQKMLERLEKENKDLRKVVLQKDEKGIHQK-----IK 273
 QY 238 AVLRDFEYSMPFIPSOFE-----LPR 258
 Db 274 KSIDMYSEVLIDLSDFDSNYNTQDHLPR 302

RESULT 7

S21801
 myosin heavy chain, neuronal [similarity] - rat
 N; Alternate names: myosin II
 N; Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S21801 #PNO013; S18134
R;Sun, W.; Chanletier, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A;Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A;Reference number: S21801; MUID:92235856; PMID:1569576
A;Accession: S21801
A;Molecule type: mRNA
A;Residues: 1-1999 <SUN>
A;Cross-references: UNIPROT:Q63731; EMBL:X62659
R;Sun, W.; Chanletier, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A;Title: A unique cellular myosin II exhibiting differential expression in the cerebellum
A;Reference number: S21801; MUID:91151356; PMID:1998509
A;Accession: PNO013
A;Molecule type: mRNA
A;Residues: 1914-1998, I' <SU2>
A;Experimental source: brain
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolyase; methylated amino acid; nucleotides
F;84-763/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;541-575/Region: actin binding #status predicted
F;553-675/Region: actin binding #status predicted
F;836-1939/Domain: coiled coil #status predicted <COI>
F;836-1276/Region: S2
F;1277-1999/Region: light meromyosin
F;125/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F;180/Binding site: ATP (lys) #status predicted
F;693,703/Active site: Cys #status predicted
F;1916/Binding site: phosphate (Ser) (covalent) #status predicted
F;1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 6.0%; Score 104.5; DB 1; Length 1999;
Best Local Similarity 26.5%; Pred. No. 24;
Matches 40; Conservative 29; Mismatches 63; Indels 19; Gaps 5;

QY 99 DEIRETLATRIHELPLSLDLDTGLHMLNCSKMLPADITQLNNIPTGQESYDNLPP 157
 ::: :::: :::: :::: :::: :::: :::: :::: :::: :::
DB 1264 ERRTVELADKKVQLQVELDNVTGLSSDSKSKLTKDFSALESQLODTQELLQEENRQK 1323
 ::: :::: :::: :::: :::: :::: :::: :::: :::: :::
QY 158 VTKSIVTNCRPVTDKIHKAYDKNKRYFEIMGEEFAFKIRTNVSHVGQLDDIRKNPR 217
 ::: :::: :::: :::: :::: :::: :::: :::: :::: :::
DB 1324 L--SLSTKLQVED-----EKNSFREQUEESEEAKHNLEKQIATLHAQVADMKKOME 1374
 ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 218 KVVCLNDINIDNHKDQTVKAVLRDYFSM 247
 ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1375 DSVGCL-----ETAEEVKRKQLQKLEG 1397
 ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 8
F71239
hypothetical protein PH0176 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: F71239
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Kino, Y.; Yamamoto, S.; Sekiguchi, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71239
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-690 <KAW>
A;Cross-references: UNIPROT:O57915; GB:AF000001; NID:g3236128; PIDN:BAA29245.1; PID:g3236128
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0176

Query Match 6.0%; Score 103. DA 2; Length 690.

	Best Local Similarity	20.7%; Pred.No.8.1;			
	Matches	66; Conservative	55; Mismatches	114; Indels	84; Gaps
QY	5	DSLYRKNILNSKFGFTSRKVPAPAHMHIDRIVMGELQMPFEEFDKTSFKHKVRHSDMQ	64		
Ddb	350	EGLRAYAVNYLGSKVVVTSATMPSLLREKLEEVGLKEL-----IPFEKVRRRYESK	399		
QY	65	PAFSYFFYLMSAVOPLNISQVFPEVDTDQSGLVSOREINTLATRIHELPLSLQDLTGLEH	124		
Ddb	400	RVRVWF-----RDLFMFEDDISIIIREKGKVLIVITNTVTRARELYEELKKRRDQVLFH	453		
QY	125	-----MLIN--CSKMLPAIDITQLNNIPTQESYYDPNPFPVTSLVNTCKRPVT	170		
Ddb	454	SRTVTRDKKEKRLVNEISSGLVA--TVQ--VEVSLDDIDYD-----TLYTEVAPID	501		
QY	171	DKI-----HKAYDKXKYRPFIMGBE---EIAFMKITNVHVVGOLDIRKNPKFV	220		
Ddb	502	SLIQRFGRVNRRGMKEGRAVYAVEGRFYLPYSKRSVEASLS-MVKLEE-ANKELDLFL	559		
QY	221	CLNNDINHKKDAQTVKAVLRDFVESMFPPSPFELPREYN-----RFLMHLEQWRAY	276		
Ddb	550	RLNDS-----PYEIWD-EYESELKKXYLERQALRIHRFRKESWLST	601		
QY	277	RD-----KLKFWTHCV	287		
Ddb	602	EDTFMSLPAIPLKENVNV	620		

RESULT 9

D90124
hypothetical protein rad3 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D90124
R:RDouglas, S.; Zaunier, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reid

```
A;Residues: 1-706 <DOU>
A;Cross-references: UNIPROT:Q98S94; GB:AF0803031; NID:g13794312; PIDN:AAK39689.1; GSPDB:IG
C;Genetics:
A;Gene: rad3
A;Map position: 3
A;Genome: nucleomorph
C;Keywords: nucleomorph

Query Match          5.9%; Score 102.5; DB 2; Length 706;
Best Local Similarity 21.9%; Pred. No. 9.1;
Matches 77; Conservative 55; Mismatches 144; Indels 75; Gaps 17;

QY      6 SLRYVVKILMSKGFTSR-KVPAMHMHID-----RIVMQELQDMPEEFDKTS----- 53
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :
Db      135 SLDVIIRSSLSKGFDFVNFKFLKPCNFYINFLKKRIKYNYIDTRKILFERNKVCPY 194
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :

QY      54 FHKVRHSDDQFAFSFYIYLMASVQPLNISQ-----VFDEVDTQSGLVSDEI----- 102
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :
Db      195 FFILNNVEDYNFLILKFQDIHHSSKNYLNPKKKIKFLPSNFES-PSALLKSKEISLS 253
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :

QY      103 -RLATRIHELPLSLQLDTGLEHMLNCMKLPADITQLANNIPTQESYY-DNPLPPTVK 160
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :
Db      254 LKTIIDSYSRLIIIVERDIIKKKIFQINVLK-----TFNNPHLKKGNFFYIKNKYKILAN 306
       ||| | | | | | | | : : : : | : | : | : | : | : | : | : | : | :

QY      161 SLVTNCKPVTDTHKA VKDNKYRFETMGEEIAFKMRTNVSHVGQL-----D 210
       ||| | | | | | | | : : | : | : | : | : | : | : | : | : | :
Db      307 YLVTKLNFQNKFHLLEYI-LNSYRFFI-----ETLLKLISKEEKYISSINYKDIFLDEN 361
       ||| | | | | | | | : : | : | : | : | : | : | : | : | : | :

QY      211 DIRKNPRKFCVLANDNIDH-----NHKDAQTVXAVLRDFYESMFFIPSQFELPREYRN 262
       ||| | | | | | | | : : | : | : | : | : | : | : | : | : | :
Db      362 FIETAPVVWFONENDNI SHYCKTYSFPFNKSLNGIKKI-----SNFLILTKEFL--NYQN 413
       ||| | | | | | | | : : | : | : | : | : | : | : | : | : | :
```

QY 263 RELHM-HELQEWRAYRDKLFWTHCVLATILMFTIFSFAEQLIAKRFKIF 312

RESULT 10

S59811
vacuolar segregation protein PEP7 - yeast (*Saccharomyces cerevisiae*)
N;Alternate names: protein D9798.11; protein YDR323c; vacuole segregation protein VAC1
C;Species: *Saccharomyces cerevisiae*
C;Date: 13-Jan-1996 #sequence revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: S59811; S59789; A41756
R;Webb, G.C.
submitted to the EMBL Data Library, March 1995
A;Reference number: S59811
A;Accession: S59811
A;Molecule type: DNA
A;Residues: 1-515 <WEB>
A;Cross-references: UNIPROT:Q92229; EMBL:U22070; NID:G736413; PIDN:AAB60290.1; PID:G736436
A;Experimental source: strain S288C
R;Du, Z.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of *S. cerevisiae* cosmid 9798.
A;Reference number: S59418
A;Accession: S59789
A;Molecule type: DNA
A;Residues: 1-515 <DUZ>
A;Cross-references: EMBL:U23517; NID:G914989; PIDN:AAB64759.1; PID:G915000; MIPS:YDR323c
A;Experimental source: strain S288C (AB972)
R;Weisman, L.S.; Wickner, W.
J. Biol. Chem. 267, 618-623, 1992
A;Title: Molecular characterization of VAC1, a gene required for vacuole inheritance and
A;Reference number: A41756; MUID:92112720; PMID:1730622
A;Accession: A41756
A;Molecule type: DNA
A;Residues: 1-365,'Y',367-515 <WEI>
A;Cross-references: GB:M80596; NID:G173156; PIDN:AAA35203.1; PID:G173157
C;Genetics:
A;Gene: SGD:PEP7; VAC1; VPS19; VPT19; VPL21
A;Cross-references: SGD:S0002731; MIPS:YDR323c
A;Map position: 4K
C;Function:
A;Description: required for vacuole segregation and vacuolar protein sorting; required
pathways

```

Query Match      5.9%; Score 102; DB 2; Length 515;
Best Local Similarity 22.8%; Pred. No. 6.6;
Matches 58; Conservative 38; Mismatches 96; Indels 62; Gaps 12

QY    28 HMP-----HMIDRVNQELQDMFPPEEFKTSFHKVRHSEDMQFAFSFYIYLSAVQPL 80
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB    281 HIPISIRLCSHCIDWL-----FIGRKFNK-----DVRMPLSGIFAKYDSMQ-- 321

QY    81 NISQVFDEVDTDQSGLVSDRIPLATRIHEL-PLSLQDLTGLEHMLNCSKMLPADITQ 139
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB    322 NISKVIDSLLPIFDLSNLKVEFAKSENSTLOPKNLDAURLKHLLNSFWLYNTLRQ 381

QY    140 L-----NNTPPQESYYDPNPFPVTKSLTVNCCKPVTDKIHKAYDKKNKYRF 185
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB    382 LLSVEPQSHLERQLQNSIKIAAAYINEKILPL-KSLPALINP---EGHKTVEDGQAEP 437

QY    186 EI-----MGEEIIAFK-----WIRTNVSHVV-QQLDDIRKNFR--KFVCLNDNDIH 228
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB    438 EVKLUSQLMIENLTTKVKELREBLMWLKQOSYLIESTIQDYKKORRLBEIVTLNKULEE 497

QY    229 NHKDAQTVKAVLRD 242
      |||   |||   |||
DB    498 LRSRIHTVQSKLGD 511

```

RESULT 11

AH1144
probable membrane protein lmo0559 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH1144
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11579669
A:Accession: AH1144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <GLA>
A:Cross-references: UNIPROT:Q8Y9G9; GB:NC_003210; PIDN:CAC98638.1; PID:g16409935; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0559

Query Match 5.9%; Score 101.5; DB 2; Length 301;
Best Local Similarity 20.2%; Pred. No. 3.7;
Matches 53; Conservative 44; Mismatches 82; Indels 83; Gaps 12;
QY 57 VRHSDM---QFASFYFYLMSAVQPLNISQVFEVDVTDQSGVLSDRITLARIHELP 113
DB 109 LKESDEVTIFOLFSSYFIMDAFPF-----YVEEMTDRTINDLKIKITTKNL----- 159
QY 114 LSLQDL-TGLEHMLINCSKMLPADITQLNNIPPTQESYVDPLPPVTKSLVTNCKPVTDK 172
DB 160 LSLSDLETGIVY-FVSASKQNAALLEQM-----K 187
QY 173 IHKAYDKNKYRFIMGEEIEAFKMI--RTNV-SHVVGQLDDIRKNPRKVCVCLNDINHDN 229
DB 188 AHMYRELNEVEKEQFDALIEARQLVEMTGLSSQILQQLSGTYNN-----ILNNLNNDT 242
QY 230 HKDAQ-----TKAVLRDFYESMFPFIPQFELPREYRNFLHMHQLQEWRAVDKLFK 283
DB 243 MKILTAISILTVTITGTGFGNMPLPL-----HNTF-----GW 278
QY 284 THCVLATLIMTIPSFFAEQLI 305
DB 279 IVTIFISVILWFLGLSILRLKIM 300

RESULT 12

S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMOF>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>

F:92/Binding site: ATP (Lys) #status predicted

Query Match 5.8%; Score 101; DB 1; Length 2663;
Best Local Similarity 20.5%; Pred. No. 62;
Matches 60; Conservative 47; Mismatches 127; Indels 58; Gaps 10;
QY 33 IDRIVWQELODFPPEFDKTSFHKVRHSDMQFASFYFYLMSAVQPLNISQVFEVD 92
DB 1972 ILEQKKEQLQURVEDVNMHKKINEMEQKKQKQEPNYLCKCEMDFOLTKKUHE---- 2027
QY 93 QSGVLSDRITLARIHELPISLQDLTGLEHMLINCSKMLPADITQLNNIPPTQESYVD 152
DB 2028 ----SLESIRIVAXERDELRIKESLKWEROQFIATLREMIARDQNHQVKEPKRLSD 2082
QY 153 PNLPPVTKSLVTNCKPVTDKIHKAYDKNKY-----REINGEEIEAFKMIKRTNVSHVG 207
DB 2083 -QQHLMESLREKCSRIKELL-KRYSEMDHDYECNLRLSLDLKEIEFHRIMKKLYLS 2140
QY 208 QLLDIRKNPRKVCVCLND-----NIDNHKDAQT----- 235
DB 2141 VYTKIEEQHE--CINKEMDFIDEVEKQKELLIKIQLQDCDVPSELRLDLKLQNWD 2198
QY 236 --VKAVLRDFYESMFPF-IPQFELPREYRNFLHMHQLQEWRAVR---DKLK 281
DB 2199 LHIEBILKDFSEFPFSIKTEFQ--QVLSNRKEMTQFLEEWLNTFRDIEKLK 2248
RESULT 13
S20596
immediate early protein - Bombyx mori nuclear polyhedrosis virus
C:Species: Bombyx mori nuclear polyhedrosis virus, BmSNPV
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: S20596
R:Ruybrechts, R.; Guarino, L.; van Brussel, M.; Vulsteke, V.
Biochim. Biophys. Acta 1129, 328-330, 1992
A:Title: Nucleotide sequence of a transactivating Bombyx mori nuclear polyhedrosis virus
A:Reference number: S20596; MUID:92162753; PMID:1536885
A:Accession: S20596
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-587 <HUY>
A:Cross-references: UNIPROT:P33245; EMBL:X58442; NID:g58714; PIDN:CAA41348.1; PID:g5871
C:Superfamily: Autographa californica nuclear polyhedrosis virus trans-activating trans
C:Keywords: transcription regulation
Query Match 5.7%; Score 99.5; DB 2; Length 587;
Best Local Similarity 23.4%; Pred. No. 12;
Matches 58; Conservative 34; Mismatches 103; Indels 53; Gaps 11;
QY 39 QELQDMFPEE-----FDKTSFHKVRHSDM-QFASFYFYLMSAVQPLNI-----SQ 84
DB 189 QBITHYFTNDFAPYLMRFDDNDYNSRFSDHMSGTGYMVFYVKSEVKPFPIIFAKYVSN 248
QY 85 VPDEVDTQSGVLSDRITLARIH-----ELPLSLQDLTGLEHMLINCSK 131
DB 249 VVVEY-TNNYMDNRVVFVDFKRFMLSYNLVKETGIEIPIHS-QDVCDETAQONCK 306
QY 132 MLPADITQLNNIPPTQESYVDPLPPVTKSLVT-----NCKPVTDKIHKAYDKNKY 183
DB 307 CHEVDVH--HTFKAALTSYFNLDMYYATTFTVTLQSLGERKCGFLSKLYEMQDKNLF 364
QY 184 REEIM-----GEEIEAFK--MIRNVSHVVGQLDDIRK-----NPRKVCVCLNDINHDN 229
DB 365 TLPIMLSKESNEIETASNNFVSPYVSQILKYSIRKVKFPDPNPKYVVDNLIWN 424
QY 230 HKDAQTVK 237
DB 425 KKSITLYK 432

RESULT 14

A61231
myosin heavy chain nonmuscle form A - human

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:46:58 ; Search time 47.0289 Seconds
(without alignment)
4012.911 Million cell updates/sec

Title: US-10-023-888-5

Perfect score: 1731

Sequence: 1 DTFADSLRYVNKILNSKFGF.....RKIFPRRIHKEASPNRIV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	490	2 Q86TQ2	Q86TQ2 homo sapien
2	1731	100.0	950	2 Q9ULJ2	Q9ULJ2 homo sapien
3	1698	98.1	921	2 Q6P9R5	Q6P9R5 mus musculu
4	1698	98.1	921	2 AAH60638	AAH60638 mus muscu
5	1290	74.5	248	2 Q9NPW9	Q9NPW9 homo sapien
6	704	40.7	132	2 Q9NV34	Q9NV34 homo sapien
7	581	33.6	384	2 Q61340	Q61340 mus musculu
8	553	31.9	717	2 Q7Q098	Q7Q098 anopheles g
9	506	29.2	666	2 Q3V553	Q3V553 drosophila
10	506	29.2	666	2 Q8SX14	Q8SX14 drosophila
11	124	7.2	1312	1 RA50_YEAST	RA50_YEAST
12	117.5	6.8	691	2 Q6FKK0	Q6FKK0 candida gla
13	113.5	6.6	384	2 Q6JUT9	Q6JUT9 pichia angu
14	113.5	6.6	384	2 AAQ74772	AAQ74772 pichia an
15	113.5	6.6	545	2 Q68215	Q68215 neisseria m
16	113.5	6.6	545	2 Q83U59	Q83U59 neisseria m
17	113.5	6.6	545	2 Q84CZ9	Q84CZ9 neisseria m
18	113.5	6.6	545	2 Q84D00	Q84D00 neisseria m
19	113.5	6.6	545	2 Q9JWW8	Q9JWW8 neisseria m
20	111	6.4	1048	2 Q26023	Q26023 plasmodium
21	111	6.4	3119	2 Q25857	Q25857 plasmodium
22	111	6.4	3119	2 Q81473	Q81473 plasmodium
23	110	6.4	1072	2 Q6FTD7	Q6FTD7 candida gla
24	108.5	6.3	480	2 Q7R8R2	Q7R8R2 plasmodium
25	106.5	6.2	483	2 Q86KM0	Q86KM0 dictyosteli
26	106.5	6.2	971	1 OPAL_ONCMA	OPAL_ONCMA
27	106	6.1	574	1 EZRA_STRPY	EZRA_STRPY
28	106	6.1	739	2 Q9M479	Q9M479 streptococc
29	106	6.1	2148	2 Q3BDJ3	Q3BDJ3 physarum po
30	105	6.1	666	2 Q90X54	Q90X54 xenopus lae
31	104.5	6.0	301	2 Q72221	Q72221 listeria mo

ALIGNMENTS

RESULT 1

ID	Q86TQ2	PRELIMINARY;	PRT;	490 AA.
AC	Q86TQ2;			
DT	01-JUN-2003 (Tremblrel. 24, Created)			
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Similar to v-maf musculoaponeurotic fibrosarcoma oncogene family,			
DE	Protein B (Avian) (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	Strausberg R.;			
RL	Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; BC042615; AAH42615.1; -			
DR	GO; GO:0005509; F:calcium ion binding; IEA.			
DR	InterPro; IPR010506; DMAP binding.			
DR	InterPro; IPR002048; EF-hand.			
DR	InterPro; IPR010983; EF-hand-like.			
DR	Pfam; PF06464; DMAP_binding; 1.			
DR	Pfam; PF00036; ehand; 1.			
DR	PROSITE; PS00018; EF_HAND; UNKNOWN_1.			
FT	NON TER 1			
SQ	SEQUENCE 490 AA; 57411 MW; 37B19FE0D1259AD2 CRC64;			
Query Match 100.0%; Score 1731; DB 2; Length 490;				
Best Local Similarity 100.0%; Pred. No. 2e-120;				
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 DTFADSLRYVNKILNSKFGFTRSRKVPAPHPMHIDRIVNQELQDMFPEEDFKTSFHKVRHS	60		
Db	163 DTFADSLRYVNKILNSKFGFTRSRKVPAPHPMHIDRIVNQELQDMFPEEDFKTSFHKVRHS	222		
QY	61 EDMQAFSYFYILMSAVPLNISQVDFEVDTDQSGVLSDREIRTLATRIHELPLSLQDLT	120		
Db	223 EDMQAFSYFYILMSAVPLNISQVDFEVDTDQSGVLSDREIRTLATRIHELPLSLQDLT	282		
QY	121 GUEHMLNCSKMLPADITQLNNIPPTQSYVDNLPVTKSLVTNCKPVTDKHKAYKDK	180		
Db	283 GUEHMLNCSKMLPADITQLNNIPPTQSYVDNLPVTKSLVTNCKPVTDKHKAYKDK	342		
QY	181 NKYRFEIINGEEBIAFKMIRTNVSHVVGQLDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL	240		
Db	343 NKYRFEIINGEEBIAFKMIRTNVSHVVGQLDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL	402		
QY	241 RDFSMSPIPSQFELPREYNRFHMHLEQEWAYRDKLFWTHCVLATIMFTIFSFF	300		
Db	403 RDFSMSPIPSQFELPREYNRFHMHLEQEWAYRDKLFWTHCVLATIMFTIFSFF	462		
QY	301 AEQLIALKRKIPPRRIHKEASPNRIV	328		

Aat03370 listeria
Q63731 rattus norv
Q81188 plasmodium
Q8k839 streptococ
Q7rrc2 plasmodium
Q813r0 bacillus ce
Q57315 pyrococcus
Q6bn4 paramecium
Q7yxz3 cryptospori
Q98894 guilladia
Q6va77 plasmodium
Aao38041 plasmodiu
Q9229 saccharomyc
Q8y599 listeria mo

```

Db 463 AEQIALKRKIFPRRIHKESPRIIV 490
|||||
RESULT 2
Q9ULL2 PRELIMINARY; PRT; 950 AA.
AC Q9ULL2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE KTAAL208 protein (Fragment).
GN Name=KIAA1208;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
DR ENBL; AB033034; BAR86522.2;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; P:cell differentiation; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR010506; DMAP binding.
DR InterPro; IPR02048; EF-hand.
DR InterPro; IPR00800; Notch region.
DR Pfam; PF06464; DMAP_binding; 1.
DR Pfam; PF00036; ehand; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT NON TER
SQ SEQUENCE 950 AA; 108785 MW; F653BDF6AEAC7503 CRC64;

Query Match 100.0%; Score 1731; DB 2; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.7e-120;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKVRS 60
DB 623 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKVRS 682
QY 61 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 683 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 742
QY 121 GLEHMLNCSXKMLPADITQLNNIPPTQESYYDNLPPVTKSLVNTCKPVTDKIKHAYKDK 180
DB 743 GLEHMLNCSXKMLPADITQLNNIPPTQESYYDNLPPVTKSLVNTCKPVTDKIKHAYKDK 802
QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKPKRFVCLNDNDHNDKDAQTVKAVL 240
DB 803 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKPKRFVCLNDNDHNDKDAQTVKAVL 862
QY 241 RDPFESMFPIPSQFELPREYRNRLHMHLEQWAEYRDKLFWTHCVLATLIMFTIFSFF 300
DB 863 RDPFESMFPIPSQFELPREYRNRLHMHLEQWAEYRDKLFWTHCVLATLIMFTIFSFF 922
QY 301 AEQIALKRKIFPRRIHKESPRIIV 328
DB 923 AEQIALKRKIFPRRIHKESPRIIV 950

RESULT 3
Q6P9R5 PRELIMINARY; PRT; 921 AA.
ID Q6P9R5

```

```

AC Q6P9R5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Udén T.B., Toshitake S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the ENBL/GenBank/DBJ databases.
DR ENBL; BC060638; AAH0638.1; -.
DR InterPro; IPR02048; EF-hand.
DR InterPro; IPR010983; EF-hand like.
DR InterPro; IPR00800; Notch region.
DR Pfam; PF00036; ehand; 1.
DR Pfam; PF00066; Notch; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 98.1%; Score 1698; DB 2; Length 921;
Best Local Similarity 97.6%; Pred. No. 1.3e-117;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKVRS 60
DB 594 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQELQDMFPPEFDKTSFHKVRS 653
QY 61 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 654 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 713
QY 121 GLEHMLNCSXKMLPADITQLNNIPPTQESYYDNLPPVTKSLVNTCKPVTDKIKHAYKDK 180
DB 714 GLEHMLNCSXKMLPADITQLNNIPPTQESYYDNLPPVTKSLVNTCKPVTDKIKHAYKDK 773
QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKPKRFVCLNDNDHNDKDAQTVKAVL 240
DB 774 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKPKRFVCLNDNDHNDKDAQTVKAVL 833
QY 241 RDPFESMFPIPSQFELPREYRNRLHMHLEQWAEYRDKLFWTHCVLATLIMFTIFSFF 300
DB 834 RDPFESMFPIPSQFELPREYRNRLHMHLEQWAEYRDKLFWTHCVLATLIMFTIFSFF 893

```

```
QY 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328
Db 894 AEQIALKRRKIFPRRIHKEASPNRIRV 921

RESULT 4
AAH60638
ID AAH60638 PRELIMINARY; PRT; 921 AA.
AC AAH60638;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.B., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060638; AAH60638.1; -
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 98.1%; Score 1698; DB 2; Length 921;
Best Local Similarity 97.6%; Pred. No. 1.3e-117;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYNKLINSKFGFTSRKVPAPHPMHIDRIYVQELQDMFPPEFDKTSFHKVRS 60
Db 594 DTFADSLRYNKLINSKFGFTSRKVPAPHPMHIDRIYVQELQDMFPPEFDKTSFHKVRS 653

QY 61 EDMQFAFSFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120
Db 654 EDMQFAFSFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHDLPLSLQDLT 713

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTCNKPVTDKHKAYKDK 180
Db 714 GLEHMLNCSKMLPADITQLNNIPPTQESYDNPVTKSLVTCNKPVTDKHKAYKDK 773

QY 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDINDHNKDAQTAVKL 240
Db 774 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDINDHNKDAQTAVKL 833

QY 241 RDFSVMFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 834 RDFSVMFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLKFTHCVLATLIMFTIFSFF 893
```

```
QY 301 AEQIALKRRKIFPRRIHKEASPNRIRV 328
Db 894 AEQIALKRRKIFPRRIHKEASPNRIRV 921

RESULT 5
Q9NFW9
ID Q9NFW9 PRELIMINARY; PRT; 248 AA.
AC Q9NFW9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp762B226 (Fragment).
GN Name=DKFZp762B226;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359588; CAB94874.1; -
DR PIR; T50618; T50618.
DR InterPro; IPR002048; EF-Hand.
DR InterPro; IPR010983; EF-Hand like.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 248 AA; 29336 MW; D4C4CA6C4505522 CRC64;

Query Match 74.5%; Score 1290; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 5.6e-88;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VFEDVTDQSGVLSDRREIRTLATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 144
Db 5 VFEDVTDQSGVLSDRREIRTLATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 64

QY 145 PTQESYDNPVTKSLVTCNKPVTDKHKAYKDKNKPVEINGEIEIAFKMIRTNVSH 204
Db 65 PTQESYDNPVTKSLVTCNKPVTDKHKAYKDKNKPVEINGEIEIAFKMIRTNVSH 124

QY 205 VVGQDDIRKNPKRFVCLNDINDHNKDAQTAVKLRFYESMFPPIPSQFELPREYRNRF 264
Db 125 VVGQDDIRKNPKRFVCLNDINDHNKDAQTAVKLRFYESMFPPIPSQFELPREYRNRF 184

QY 265 LHMHELOEWAYRDKLKFTHCVLATLIMFTIFSFFASQIALKRRKIFPRRIHKEASPN 324
Db 185 LHMHELOEWAYRDKLKFTHCVLATLIMFTIFSFFASQIALKRRKIFPRRIHKEASPN 244

QY 325 RIRV 328
Db 245 RIRV 248

RESULT 6
Q9NV34
ID Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10959.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
```

```

RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshina A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Okamoto S.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RA "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK001821; BAA91926.1; -.
SQ SEQUENCE 132 AA; 16083 MW; 92B5DEB4E062AD07 CRC64;

Query Match 40.7%; Score 704; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 9.9e-45;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MIRTNSHVYQGLDDIRKPRKEVCLNDNDHNDKDAQTVKAVLRDFYSEMPPIPSQFEL 256
DB 1 MIRTNSHVYQGLDDIRKPRKEVCLNDNDHNDKDAQTVKAVLRDFYSEMPPIPSQFEL 60

QY 257 PRYRNFRLMHELOEWRAYRDXLKFVTHCVLATLIMFTSFPAQIALKRRKIPRRR 316
DB 61 PRYRNFRLMHELOEWRAYRDXLKFVTHCVLATLIMFTSFPAQIALKRRKIPRRR 120

QY 317 IHKEASPNRIRV 328
DB 121 IHKEASPNRIRV 132

RESULT 7
Q61340 Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01; Created)
DT 01-NOV-1996 (TrEMBLrel. 01; Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26; Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN Name=MafB;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX Cordes S.P., Barsh G.S.;
RA "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1025-1034 (1994).
DR EMBL; L36434; AAA65688.1; -.

```

```

DR PIR; I49528; I49528.
DR MGD; MGI:104555; MafB.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0045944; P:positive regulation of transcription from P...; IDA.
DR GO; GO:0007585; P:respiratory gaseous exchange; IMP.
DR GO; GO:0007379; P:segment specification; IMP.
DR InterPro; IPR010506; DMAP_binding.
DR InterPro; IPR02048; EF-hand.
DR InterPro; IPR010983; EF_Hand_Like.
DR Pfam; PF06464; DMAP_binding; 1.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 384 AA; 43753 MW; 51F473C8807A7E55 CRC64;

Query Match 33.6%; Score 581; DB 2; Length 384;
Best Local Similarity 97.4%; Pred. No. 5.3e-35;
Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMHIDRIYVQELQDMFPPEFDTSPFKVRS 60
DB 267 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMHIDRIYVQELQDMFPPEFDTSPFKVRS 326

QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHELPLS 115
DB 327 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHDPLT 381

RESULT 8
Q7Q098 Q7Q098 PRELIMINARY; PRT; 717 AA.
AC Q7Q098;
DT 01-MAR-2004 (TrEMBLrel. 26; Created)
DT 01-MAR-2004 (TrEMBLrel. 26; Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26; Last annotation update)
DE AGCP9986 (Fragment).
GN Name=agCG49111; ORFNames=ENSAANG00000017589;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
CX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAH01008986; EAA00384.1; -.
FT NON_TER 1
SQ SEQUENCE 717 AA; 82794 MW; 0BA7D2B957E48250 CRC64;

Query Match 31.9%; Score 553; DB 2; Length 717;
Best Local Similarity 39.6%; Pred. No. 1.4e-32;
Matches 112; Conservative 57; Mismatches 104; Indels 10; Gaps 5;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMHIDRIYVQELQDMFPPEFDTSPFKVRS 60
DB 399 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHPMHIDRIYVQELQDMFPPEFDTSPFKVRS 458

QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHELPLSLODIT 120
DB 459 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHELPLSLODIT 518

QY 121 GLEHMLNCS---KMLPADITQ--LNNIPPT--QESYYDNLPPVTKSLVNTCKPVTDKI 173
DB 519 YFEVWVNCSLQNNFLPLESHQEQRQVYPLTVYVEREDSTIPVTRSLVVGCAELADM 578

```


QY 174 HKAYDKXKXRYFEMGEEI--AFKJRTNVSHVVGOLDIRKXPRKFCVCLNDNI-DHNH 230
 Db 579 RANFGKVPYRYHYSAKTYISNFKMTSNTITQVVDALDBLRKTPKFKPCINDNLSDDRP 638
 QY 231 KDAQTVKAVLDRDFVESMFPIPSQELPREYVNRFLMHHEIQEW 273
 Db 639 DDNLQIGALLEBDFYLSLFPARSSPELESTYVNRFRQYDDYRSV 681

RESULT 9
 Q9V553 PRELIMINARY; PRT; 666 AA.

AC Q9V553;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE CGS027-PA.

GN ORFNames=CGS027;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]

SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champ M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodrager, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]

SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Swirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.,

RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
 melanogaster euchromatic genome sequence";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]

SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.,
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomics perspective";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]

SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Beriman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.,
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]

SEQUENCE FROM N.A.
 RP FLYBASE;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]

SEQUENCE FROM N.A.
 RG FLYBASE;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AR003834; RAF58967.2; -;
 DR Intact; Q9V553; -;
 DR Flybase; FBgn0033392; CGS027;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0030154; P:cell differentiation; IEA.
 DR InterPro; IPR000800; Notch_region.
 DR Pfam; PF00066; Notch; 1.
 SQ SEQUENCE 666 AA; 77731 MW; EA232EC57C754FF6 CRC64;

Query Match 29.2%; Score 506; DB 2; Length 666;
 Best Local Similarity 33.6%; Pred. No. 4e-29; Mismatches 133; Indels 10; Gaps 5;
 Matches 108; Conservative 70;

QY 1 DTAFDSLRYVKNILSKFQTSRKVPFAHMPHMDIRVIMQELQDMFPPEFDTKTSFEKVRHS 60
 Db 348 DIYSHSLIATNNLLNRAVGFKAHVLAHVGFIDKDIVEMQSRFHOQILDTAHQRFAP 407

QY 61 EMQAFASYFYILMSAVQPLNISQVDFDQSVLSDEIRLTATRIHELPLSLQDLT 120
 Db 408 TDLQAFAYFYSILMSETKMSVEEIFDFDQSVLSDEIRLTATRIHELPLSLQDLT 467

QY 121 GLEHMLNCSK---MLPADITQLNIPPTQBSYVDNLPVTKSLVNCRPVTDKIHKA 176
 Db 468 YEEVVQCNTRNLGMHLKVDVTEHSTL--VVERYEDSNLPTITRDLVRCPLIALAALAN 525

QY 177 YKDKNKYRFEINGE--EIEAFQKHTNVSHVVGQDDIRKXPRKFCVCLNDNI-DHNH-KDA 233
 Db 526 FAVRPKNYFHVSKTSHSNFMMLTSLNLTVEVDESILDRLRNRPKFCINDNLDNRGDN 585

QY 234 QTVKAVLDRDFVESMFPIPSQELPREYVNRFLMHHEIQEWAYRDKLFWTHCVLATLIM 293
 Db 586 EWMRHLLDFYLSLFPARSSPELESTYVNRFRQYDDYRSV 644

QY 294 FTIFSFFAEQLIAKXKIFPR 314
 Db 645 VCLLRFMCHHKAKLVRCVQR 665

RESULT 10
 Q8SX14

```

ID QBSX14 PRELIMINARY; PRT; 666 AA.
AC QBSX14;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE R235033p.
GN ORFNames=CG8027;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.B., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089618; AAL90356.1; -
DR FlyBase; FBgn0033392; CG8027.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_region.
DR Pfam; PF00066; Notch; 1.
SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;

Query Match 29.2%; Score 506; DB 2; Length 666;
Best Local Similarity 33.6%; Pred. No. 4e-29;
Matches 108; Conservative

Qy 1 DTPADSLRYNKILSKFGTGRKVPAMPHMDIRVIMQELQDMPEEEDKTSFKVHRIS 60
Db 348 DIYSHSLIATNMLNIRAYGFAKHLVAHVGLDKDIVEAMQRHFQQLDIAHQFRAP 407

Qy 61 EDQAFASYFYVLMASVQPLNISQVDFEVDTPDQSVLSREITLATRIHEPLSLQDLT 120
Db 408 TDQYAFASYFLMSETKVMSVEIFEPTDQSATWSDEVTFTLRIYQPLDWSNR 467

Qy 121 GLEHMLNCSK---MLPADITQLNNIPPTQSYDNPVPTKSLVITNCKPVTDKIHA 176
Db 468 YFEVQVQCTRNGLMHLKVDITVEHSTL--VYERYEDSNLPTITRDVWRCPLIAEALAN 525

Qy 177 YKDKNYRFEIMGE--EIIAFKIRTNVSHVQGLDDIRKPKFKVCLNDNDINH-KDA 233
Db 526 FAVRPKYNFVSPRTSHSNFMMLTSLNTEVSDRLRKNPKFNCINDNDANRGEDN 585

Qy 234 QTVKAVLRDFESMFPISPFQELPREYRNRLHMLQELQEWRAVYDKLKFVTHCVLATLM 293
Db 586 EMVEHLLDFYLSFFPERSKFELPPQYRNRFESWRDQWRK-RKRAVLVIGVGVSLLV 644

Qy 294 FTIPSPFAEQIALKRIKIFR 314
Db 645 VCLLRFMCHKAKLVRECVQR 665

RESULT 11
RA50 YEAST
ID RA50 YEAST STANDARD; PRT; 1312 AA.
AC P12753;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA repair protein RAD50 (153 kDa protein).
GN Name=RAD50; OrderedLocusNames=YNL250W; ORFNames=N0872;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=RE821;
RX MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions.";
RL Genetics 122:47-57(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SU11 from
RT the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site.
CC -!- SUBUNIT: Forms a complex with MRE11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X14814; CAA32919.1; -
DR EMBL; X96722; CAA65494.1; -
DR EMBL; Z71526; CAA96157.1; -
DR PIR; S05808; BWSYDL.
DR Germonline; 143256; -.
DR SGD; S0005194; RAD50.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003690; F:double-stranded DNA binding; IEA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0006303; P:double-strand break repair via nonhomologous...; IMP.
DR GO; GO:0000722; P:telomerase-independent telomere maintenance; IMP.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR007517; Rad50_zn_hook.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR TIGRFAMs; TIGR00606; rad50; 1.
KW ATP-binding; Coiled coil; DNA repair; Hydrolase; Meiosis.
FT NP_BIND 34 41 ATP (Potential).
FT DOMAIN 185 347 Coiled coil (Potential).
FT DOMAIN 403 558 Coiled coil (Potential).
FT DOMAIN 617 672 Coiled coil (Potential).
FT DOMAIN 734 1108 Coiled coil (Potential).
SQ SEQUENCE 1312 AA; 152568 MW; 58A0AA173AC5677E CRC64;

Query Match 7.2%; Score 124; DB 1; Length 1312;
Best Local Similarity 21.3%; Pred. No. 2.5;
Matches 68; Conservative 65; Mismatches 136; Indels 60; Gaps 14;

Qy 5 DSLRYVNVKILSKFGTGRKVPAMPHMDIR-----IVMQELQ---DMFPEEFD--KTSF 54
Db 189 DNLKSIKDMSVDIKLJKQSV-EHLKLDKRSKAMKLNHQLQTKIDQYNEEVSEISQL 247

Qy 55 HKVHSEDMQAFSY-FYVLMASVQPLNISQVDFEVDTPDQSVLSREITLATRIHELP 113
Db 248 NEITEKSDPKFKSNQDFQKILSKVENLKNL-----SISD-QVKRLNSIDILD 296

Qy 114 LSLQDLTGLHMLNCSKMLPADITQLNNIPPTQSYD--PNLPPVTKSLVITNCKPVD 171
Db 297 LSKPD---LQNLNLFKVLMDKKNQRLDETDISSLKDRQSSLSLSLIRRQGL-- 351

Qy 172 KHKAYKDKNKYRFEIMGEIEAFKIRTNVSHV-----VQQLDIRKKNPKFVC--LNDN 225

```

```
Db 352 EAGKETEKRNHLSSLKE---AFQHKFQGLSNLSDMAQVNHMSQFKAFISQDLTDT 408
Qy 226 IDNHKDAQTAVLRDFYFSMFPISQFELPRYRNRLH- 266
Db 409 IDQFAKIQIKETNLSDIKSITVDSONLEVNKDRSKLIHDSBELAELKLSFKSLSTQD 468
Qy 267 --MEHLQWRAYRDKLKF 283
Db 469 SLNHELENLTKYKELQSW 487

RESULT 12
Q6FKKO PRELIMINARY; PRT; 691 AA.
AC Q6FKKO;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Similar to tr|Q12033 Saccharomyces cerevisiae YOR275c.
DE ORFNames:CAGJ0L109569;
GN Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
CX NCBI_TaxID=5478;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG GENOLEVURES;
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckrich J.M., Beyne E., Bleykastein C.,
RA Boissame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Faithhead C., Ferry-Dumazet H., Groppi A.,
RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Ozias S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potter S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekalia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.H.;
RA "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380958; CAG62218.1; -.
SQ SEQUENCE 691 AA; 80473 MW; 12D26888D8BABC CRC64;

Query Match 6.8%; Score 117.5; DB 2; Length 691;
Best Local Similarity 19.4%; Pred. No. 3.4;
Matches 61; Conservative 54; Mismatches 128; Indels 71; Gaps 13;

Qy 23 RKVPAH-----MPHMIDRIVMOELQDMPEEFDKTSFHKVHSDMOPAFSY-- 69
Db 305 KHVESHLOFSDFKLYKLDKMDKLLVDPDTEIAPSLMDQLFNLLFIIMDYGSYNE 364
Qy 70 ---FYLLMSAVQPLNISQV-----FDEVDTQSGVLSDRIRTLATRIHELPLSLQD 118
Db 365 KQDIYVLQHFVEPINLNIQLDELASFFEID-PQHIWNAHNKTSFVTR-KELDIVGQS 422
Qy 119 LTGHEHLNCSKMLPADITQLNNIPPTQSYDPPNLPVTKS-----LVTNCKPVT 170
Db 423 FGDLESNAINIE-----TQLANI-----TSYLDLEIKTYEEDKQQGVWTVTDARDVT 471
Qy 171 DKIHKAYDKNKY-----RFEIMGEEIAFKMIRTNVSHVVGQLDD-----I 212
Db 472 KEFYEKVEKLQYLEIGRNINLELQELFASIDSKLSVTSRTSKAHLTAQYNDPFLNKTEAL 531
Qy 213 RKNPRKFVCLNDNDHDKDAQTAVLRDFYFSMFPISQFELPRYRNRLHMEHLOE 272
Db 532 KKRERFI---SEIEKSYENRIISKLIM-YKETDDIPGTISERIEIFDNIYSKH---- 583
Qy 273 WRAYRDKLKFWTHC 286
```

```
Db 584 MKVFAVDMKYIEDC 597

RESULT 13
Q6JUT9 PRELIMINARY; PRT; 384 AA.
AC Q6JUT9;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE PDD4p.
GN Name=PDD4;
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetales; Pichia.
CX NCBI_TaxID=4905;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS4732;
RC Monastyrsky I., Kiel J.A.K.W., Veenhuis M.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY304543; AAQ74772.1; -.
SQ SEQUENCE 384 AA; 45251 MW; E294618BBCE6AC12 CRC64;

Query Match 6.6%; Score 113.5; DB 2; Length 384;
Best Local Similarity 20.5%; Pred. No. 3.2;
Matches 69; Conservative 62; Mismatches 130; Indels 75; Gaps 15;

Qy 5 DSLRYVYNKILSKFGFTGRKVPAPHPHMIDRIVMOEL--QDMFPEEFDKTSFHKVHSD 62
Db 58 NAAQVYNQKLSAK-GYFSNK-----PNKLRQLLLSLDSSQLIPENAEFNFEISEKVYND 111
Qy 63 MQFASYSFYLLMSAVQPLNISQVDE-----VDTQSGVLSDRIRTLATRIHELPLSL 116
Db 112 RN-VNNIYSLNSAE---ISKQFRETALKKQADKADIELLKNEVKLNKKVDEKERNI 167
Qy 117 QDLTGLEHMLINC-SKMLPADITQLNNIPPTQSYDPPNLPVTKSLVTKNCKPVTDKIHK 175
Db 168 OSIQ-LDKIKSLDQAKHYCAMLNNVNAKKEQERTFLYABEVKEELARNEMEV-DRLES 225
Qy 176 AYKDKNKRFEIMGEEIAFK-----MIRTN-----VSHVVGQLDDIRKNPRK 218
Db 226 RLNNMNNKRASPRDDAAEAPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 285
Qy 219 FVCLNDN-----IDHNKDAQTAVLRDFYFSM-----FPIPS 252
Db 286 FQGLNSNSKVPSTPIPTDSELLAIDHNFEDLQKYDLILFDLLKPLNGSIDDTFKEKNN 345
Qy 253 QFELPRE---YRNRLHME-----LQEWYR 277
Db 346 QTDKQEIATSLNRLSEMEQYERVLATWEQWKTYR 381

RESULT 14
AAQ74772 PRELIMINARY; PRT; 384 AA.
AC AAQ74772;
DT 01-JUN-2004 (TReMBLrel. 27, Created)
DT 01-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE PDD4p.
GN PDD4;
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetales; Pichia.
CX NCBI_TaxID=4905;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS4732;
RC Monastyrsky I., Kiel J.A.K.W., Veenhuis M.;
RL "The H. polymorpha PDD4 gene product is required for selective
RT peroxisome degradation.";
RT
```

